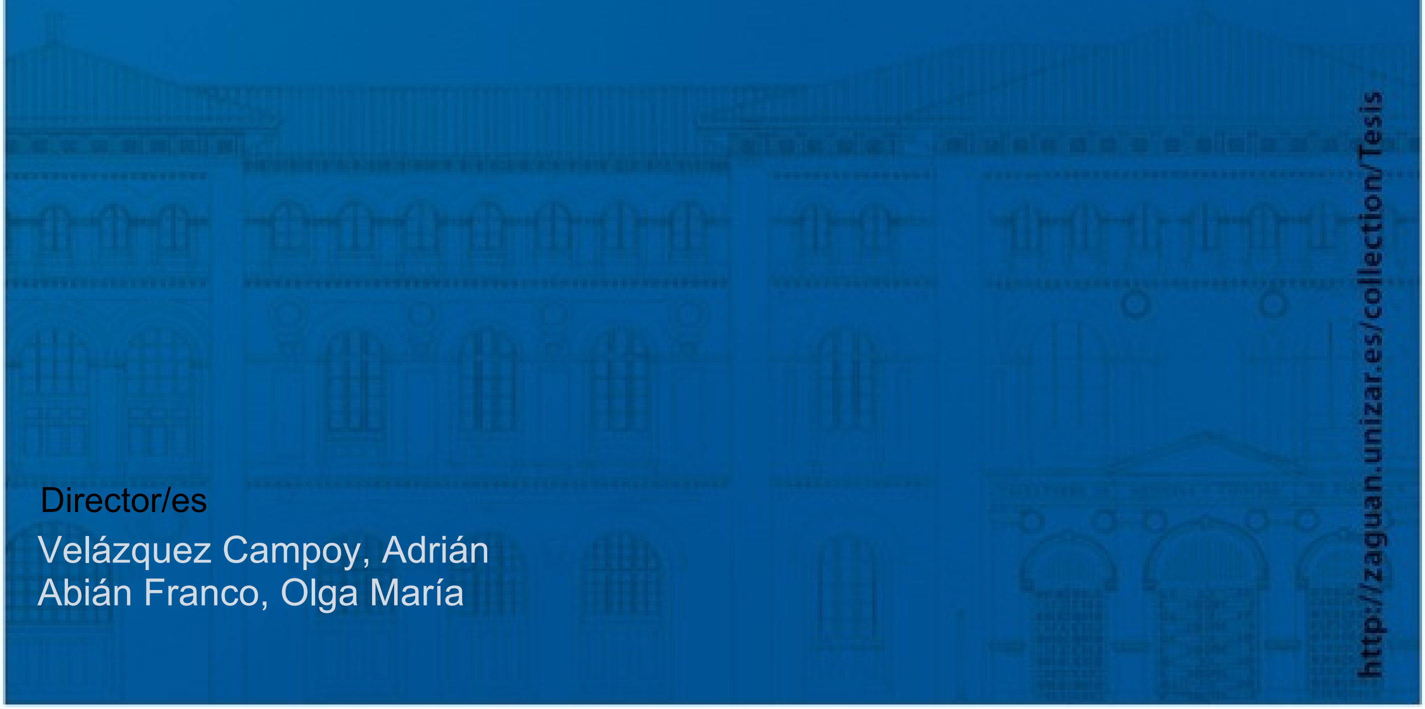
TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2024

92

Ana Jiménez Alesanco

Targeting BFT-3 and NUPR1: novel approaches for drug discovery in colorectal and pancreatic cancer therapies



Director/es

Velázquez Campoy, Adrián
Abián Franco, Olga María

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

All Rights Reserved

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Tesis Doctoral

TARGETING BFT-3 AND NUPR1: NOVEL APPROACHES FOR DRUG DISCOVERY IN COLORECTAL AND PANCREATIC CANCER THERAPIES

Autor

Ana Jiménez Alesanco

Director/es

Velázquez Campoy, Adrián
Abián Franco, Olga María

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Bioquímica y Biología Molecular

2023

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

DOCTORAL THESIS

Targeting BFT-3 and NUPR1: novel approaches for drug discovery in colorectal and pancreatic cancer therapies

MEMORIA PRESENTADA POR

ANA JIMÉNEZ ALESANCO

DIRECTORES

ADRIÁN VELÁZQUEZ CAMPOY
OLGA ABIÁN FRANCO

UNIVERSIDAD DE ZARAGOZA
2023

Memoria de Tesis Doctoral presentada por **ANA JIMÉNEZ ALESANCO** para optar al Grado de Doctora con Mención Internacional por la Universidad de Zaragoza.

Dr. **ADRIÁN VELÁZQUEZ CAMPOY**, Profesor titular del Departamento de Bioquímica y Biología Molecular de la Universidad de Zaragoza e Investigador de la Fundación Agencia Aragonesa para la Investigación y el Desarrollo (ARAID),

Dra. **OLGA ABIÁN FRANCO**, Profesora titular del Departamento de Bioquímica y Biología Molecular de la Universidad de Zaragoza e Investigadora del Instituto de Investigación Sanitaria de Aragón (IIS),

CERTIFICAN:

Que la Tesis Doctoral "**TARGETING BFT-3 AND NUPR1: NOVEL APPROACHES FOR DRUG DISCOVERY IN COLORECTAL AND PANCREATIC CANCER THERAPIES**" ha sido realizada por **ANA JIMÉNEZ ALESANCO** dentro del Programa de Doctorado en Bioquímica y Biología Molecular, en el Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), el Departamento de Bioquímica y Biología Molecular (Facultad de Ciencias), y en colaboración con el Departamento de Microbiología (Facultad de Medicina) de la Universidad de Zaragoza, y el Centro de Investigación Biomédica de Aragón (CIBA) bajo su dirección. Además, que ha realizado una estancia de 3 meses en el Grupo de Cáncer Pancreático del Cancer Research Center of Marseille (CRCM), en Marsella (Francia), y que reúne las condiciones requeridas para optar al Grado de Doctora con Mención Internacional por la Universidad de Zaragoza.

Zaragoza, 15 de septiembre

Adrián Velázquez Campoy
Director

Olga Abián Franco
Codirectora y tutora

Ana Jiménez Alesanco
Doctoranda

La realización de esta Tesis Doctoral ha sido posible gracias a la siguiente financiación:

- Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033/) y por FEDER “Una manera de hacer Europa” (PID2021-127296OB-I00).
- Diputación General de Aragón: contratación de personal investigador predoctoral en formación para el período 2019-2023 y subvención de fomento de la movilidad de personal investigador predoctoral en formación para el año 2023. Número de expediente MVE_05_23.
- Fondo de Investigaciones Sanitarias del Instituto de Salud Carlos III y la Unión Europea (ERDF/ESF, “Investing in your future”) (PI15/00663 and PI18/00349 y PI21/00394)
- Programa del Miguel Servet del Instituto de Salud Carlos III (CPII13/00017).
- Diputación General de Aragón (Grupo Protein Targets and Bioactive Compounds E45_17R y E45_23R, y Grupo de Investigación Traslacional en Patología Digestiva B25_17R y B25_23R).
- Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd).
- Plan Complementario de Biotecnología Aplicada a Salud (Línea de Actuación 3) del Ministerio de Ciencia e Innovación y el Gobierno de Aragón, con Fondos NextGenerationEU y Plan de Recuperación, Transformación y Resiliencia (PRTR-C17.I1).
- Fundación ARAID.
- Instituto de Investigación Sanitaria Aragón (IIS Aragón).
- Instituto Aragonés de Ciencias de la Salud (IACS).

ACKNOWLEDGEMENTS

Estos años, especialmente los dos últimos, no han sido nada fáciles para mí. Quien bien me conoce es consciente de ello. Soy sincera si más de una vez pensé que este momento no podría llegar, que no sería capaz. Si estoy aquí es sin duda gracias a toda la gente que nunca ha dejado de confiar en mí, incluso cuando yo no lo hacía. Desde hace mucho tiempo me he estado visualizando escribiendo los agradecimientos de esta Tesis. Sí, es así. Llevo mucho tiempo queriéndolo hacer porque lo veía como un punto en el que por fin poder abrirme y dar las gracias. Las gracias de corazón. Y sí, me voy a explayar. Pero es algo que mi *patata* me pide hacer.

Antes de empezar a ponerme sensiblona, que espero no hacerlo en exceso (pero lo voy a hacer porque ya me conocéis) quisiera agradecer por orden a quien ha hecho posible oficialmente que haya llevado a cabo esta Tesis Doctoral. En primer lugar, esta Tesis no hubiera sido posible sin la subvención para contratación de personal investigador predoctoral en formación obtenida de la Diputación General de Aragón. Espero haberla aprovechado al máximo y haber aportado mi pequeño granito de arena al mundo de la investigación. Por supuesto, gracias a mis directores, Olga y Adrián, por la enorme oportunidad que un día allá por el 2018 me disteis, por permitirme entrar en el mundo de la ciencia, por la confianza y libertad plena como investigadora que habéis tenido en mí, y por todo vuestro tiempo invertido. Me llevo de vosotros una incontable cantidad de conocimientos aprendidos, tanto profesionales como personales, y siempre estaré agradecida por permitirme colaborar y participar en proyectos científicos tan diferentes. También gracias por haberme dado ánimos en incontables ocasiones, y por haberme enseñado que lo positivo atrae a lo positivo. En segundo lugar, gracias a todas esas personas que han formado y forman parte del BIFI, y que han hecho que me sienta parte viva de él. Sonia, gracias por decirme tantas veces que soy una pesada. Lo soy, sí. Y aunque a veces me haya enfurruñado por algo que me habías dicho (la gran mayoría de veces teniendo tú razón…la mayoría ¿eh?), tengo que reconocer que, si alguien tiene que poner *orden en la sala* para que todo siga funcionando como funciona, esa persona sin duda eres tú. Gracias por todo lo que me has enseñado estos años y por todos los salseos que hemos compartido. Gracias también a Hajar, Marta y Paula por tomarnos el relevo a David y a mí, y al resto de Investigadores Principales del BIFI, que siempre me han ayudado y aportado soluciones cuando he ido a buscarlos con alguna cuestión.

Una de las mejores cosas que me ha dado este trabajo es, sin duda, la gente. Cuando llegas a un sitio nuevo, el miedo es un sentimiento muy presente (y como sabéis soy sensiblona y pesada, pero también miedica). Gracias por el acogimiento que me disteis desde el primer momento. Recuerdo que a la semana y media de estar en el BIFI ya nos fuimos unas 15-20 personas de casa rural al Pirineo. ¡Qué maravilla! Y es que eso fue el comienzo de una bonita amistad que sé que mantendré para toda la vida. Pero antes de irnos al ocio (sensiblona/pesada/miedica/ociosa) y contar lo que se puede contar del espacio exterior al BIFI, quiero hablar de lo que ha ocurrido dentro de estas cuatro paredes (nota importante: este espacio exterior existe, aunque a veces haya dudas por las horas empleadas en el mismo). Dentro del BIFI, no solo me habéis apoyado cuando no salía un experimento, enseñado cómo utilizar un aparato o cómo hacer un ensayo, me habéis apoyado cuando ya no tenía fuerzas para continuar y me habéis cuidado como grandes amigos. David, creo que una de las cosas más importantes de tu trabajo, es sin duda poder realizarlo de la manera más cómoda posible con la gente que tienes más próxima. Gracias por ser el compañero de grupo que has sido, tanto en los momentos de caos, que no han sido pocos, como en los momentos de orden. Y que un compañero con el que pasas tantas horas se convierta en un gran amigo no siempre es fácil. Me has apoyado, enseñado, ayudado y salvado el pellejo en alguna que otra *mucha* ocasión. Y aunque hemos vivido momentos duros juntos, hemos reído y lo hemos pasado en grande haciendo ciencia (o más bien siensia, en la que SÍ, con todo se puede jugar), y esto es algo que siempre recordaré con infinito cariño. Tampoco puedo olvidarme de los bailoteos que nos hemos pegado, que no han sido pocos, y brindo por los que sé que vendrán. Andrés, ay mi André. Dame un cortau, pero esta vez largo de café, que cuando esto acabe quiero celebrarlo de la mano contigo en algún bar nuevo que me han dicho que han abierto por ahí. Igual no lo conoces, tranqui yo te llevo. Si algo puedo decir de ti es que, a no ser que sea jueves, siempre estás ahí. Me has enseñado a crecer como persona, a creer en mí y también, muy importante, a reírme de mi misma y de los problemas de la vida. La vida unida al drama es mucho más divertida. Gracias por estar en mi vida, por ser un apoyo incondicional y por no odiarme junto con David por ser *normi* y no entender vuestras cositas. Bea, eres una de las personas que más cerca siento a pesar de los malditos 1.326 km de distancia que hay entre las dos, y cada vez que hablamos sé que puedo hablarte desde el corazón y la sinceridad, sé que le estoy hablando a una amiga de las de verdad. Además, a día de hoy no te puedes separar de mí, te recuerdo que seguimos casadas garajealmente hablando, con un testigo precintado. Y sí, estoy hablando de tu bici, que igual estás leyendo esto en modo *chill* y ya sabes... aissss, nunca dejes de llamarme para ponernos al día durante 2 h que pasan volando, ni de mandarme abracicos por mensaje. Ernesto, no me faltes nunca. Dame muchos más momentos de a ti, a ti, a ti. Echo mucho de menos aquellas temporadas en las que nos visitabas en el BIFI, pero también los momentos en el estanque de la city marujeando de la vida. Aun así, te tengo muy presente, no solo porque tengo la suerte de que vienes a verme mil veces y tienes tu propia habitación en mi casa, sino porque al igual que Bea, estás ahí día a día. Sigue guiando la dirección del camino correcto con tus ojos, por favor y gracias. Polanco, mi super partner. El pequeño del BIFI que de repente… ups, ya no es tan pequeño. Parece que fue ayer cuando llegaste, siempre con tus ganas de hacer el bien por los demás y ayudar. Y también por trolear, no nos engañemos. Compañeros como tu hay pocos, pero es que además hiciste que cada tarde fuera una forma de superarnos a nosotros mismos. Ojalá pronto volver contigo al campo de batalla, amigo.

Gracias a mis BIFIsaurios, que fueron referentes para mí y sigo teniendo la suerte de poder contar con ellos. Pablete, solo te diré mimimimimimi. Benditas todas las cabezonerías, quesos y cervezas que hemos compartido, amigo. Pepe, gracias por seguir actualizándome tu vida con 60 mensajes seguidos o audios de 22’23’’ (*juroqueestohaocurrido*), y preocupándote por mí. Ojalá más historias fascinantes de las tuyas, relacionadas por supuesto con comida y sus efectos. Mati, eres la viva imagen de que unas ideas locas no pueden no ir asociadas a un pelo loco. Sigue enredando mi vida cada vez que nos juntamos y… ¡qué viva el curly! María C., tienes un corazón de oro. Siempre con una sonrisa enorme que no debes perder nunca, pase lo que pase. Son incontables los consejos que me diste en mis inicios. Pabliño, gracias por toda la sabiduría que pummmmm, de repente sueltas. Y por las visitas al CIBA para tomarnos un cafecico que tanta energía me daba cuando me sentía sola trabajando. Sonia H., has estado ahí cuando necesitaba a alguien que me entendiera, y sé que seguirás dándome esos abracitos tan buenos. Violeta, gracias por tu gran apoyo dentro del grupo y por las conversaciones en las que al final sacábamos algo de luz entre tanta negrura. Víctor, eres un claro ejemplo de calma y paz en el trabajo, algo que como terremoto que soy, te agradezco que hayas aportado en todos día a día. Ánimo en esta última etapa, que en nada lo celebraremos también. Ana G., gracias por ser apoyo en tantas dudas existenciales y por supuesto científicas. Juanito, tienes el poder de hacer creer a la gente cualquier cosa incluyendo cuestionarios al taxista de turno. Gracias por los infinitos consejos que me has dado durante estos años. Laura, mis otras dos manos en el laboratorio que tanto me ayudaron con momentos duros de trabajo y que tanto cariño que me daban en esos grandes abrazos. Juanjo, el BIFIsaurio por excelencia que seguía al pie del cañón, apoyando, compartiendo momentos llenos de cariño. Bea J., mi compañera de los inicios. Estuviste en un momento en el que David y yo no nos podremos olvidar nunca de toda tu ayuda. Sandra S., siempre tan atenta por los demás y llenando las quedadas de humor y alegría. Javier M., también me podrías haber tachado por ser *normi*, pero no lo hiciste. Ha sido un placer haber trabajado contigo. Helena, eres un gran ejemplo de superación y compañerismo. Vaya cantidad de BIFIsaurios…menos mal que han ido llegando las nuevas generaciones para aportar algo de frescura a estos quejicas pesados que estamos acabando. Sandra, Carlos, Irene, Ángela, Alejandra… habéis entendido lo que este trabajo significa y sé que seguiréis cuidando el buen ambiente que siempre hemos tenido los doctorandos del BIFI. Mucho ánimo que, aunque lo veáis lejano, acaba llegando. Y sorry, pero os seguiré dando mal para que vayamos al Tony. Asociado al trabajo no puedo olvidarme de mi *quasi*Dra Marta. Gracias Andrés por poner en mi vida a una persona tan increíble y que me hace sentir como en casa. Tranquila, que no te lo voy a robar, aunque te lo repita una y otra vez. Gracias por dejar que tenga mi propia terraza-llorería personal, por todas las confesiones que tanto me ayudan a crecer personalmente y por el apoyo incondicional que me das.

Sigo teniendo que agradecer a la gente que me ha acompañado y asesorado en mi trabajo, pero ahora me tengo que ir un poquito más lejos. A mi Marsella, *ma deuxième maison*. Sin duda fue un momento clave en el final de esta Tesis. Un momento en el que tu cabeza hace “clic”. Un “clic” que muchas veces no sabemos interpretar, pero en el que el aire fresco amigos, aseguro que ayuda. ¡Y tanto que ayuda! Durante mis meses de estancia allí por supuesto aprendí muchísimo como investigadora, pero creo que aun más, si cabe, como persona. Patricia, no solo me has demostrado tu gran valor como investigadora, sino que gracias a ti he descubierto pequeños rincones y sitios de comida increíbles en Marsella. Tienes un potencial admirable en ciencia, pero también un tremendo corazón como persona. Siempre te estaré agradecida por aquellos meses en Marsella, y sé que habrá muchísimas ocasiones en las que desee ir a veros a ti y a Pepe. Dale un besito de mi parte. *Merci aussi à tous les collègues du Centre de Cancérologie de Marseille pour m'avoir accueillie comme une membre du groupe, pour les afterworks sur la plage et pour avoir été patients avec mon français, avec lequel vous m'avez beaucoup aidée*. Y gracias a mi familia fuera del laboratorio en Marsella. Me considero una gran afortunada de la oportunidad que me dio la vida de, en tan poco tiempo conocer a gente tan increíble en Marsella. Julianito, sin duda te debo buena parte de esa experiencia. *Mon petit frère français et le melleur collocateur sans doubte*. Irene, mi sevillana-francesa preferida, gracias por escucharme y hacerme la vida infinitamente más fácil allí. Ghizlaine, *tu as un cœur d'or, continue à briller. Merci à tous les partenaires du Shamrock. Je reviendrai bientôt et je sais où vous trouver, mes amis.*

Como he comentado antes, para mí ha sido un orgullo que Olga y Adrián me dieran la oportunidad de colaborar en otros proyectos. El que más quebraderos de cabeza, pero que más satisfacción me ha dado a la larga sin duda ha sido el trabajo con mis queridos gusanicos en el Centro de Investigación Biomédica de Aragón. Esto no hubiera sido posible sin la confianza de José Antonio A. y Santiago R. Gracias por permitirme trabajar y aprender tanto con vosotros. Y especialmente gracias a Marta G. Recuerdo con mucho cariño todas esas horas iniciales en las que de repente teníamos a estos bichines delante de nosotras, y pincharles fue una auténtica odisea. Gracias a tu apoyo, supe llevarlo de la mejor manera que se puede hacer de todo en esta vida. Con el pecho *palante* y buen humor. Y gracias Lara por todo el apoyo y ayuda recibidos.

Por supuesto algo he tenido que hacer con mi vida aparte del trabajo (ocio, ocio, ocio), y durante todos estos años estoy orgullosa de tener a la gente que tengo a mi lado. Abutardas, los de toda la vida, los del pueblo, los de casa. Cada uno literalmente, diferente. Pero eso nos hace únicos, y ha hecho que desde pequeños hayamos vivido millones de aventuras, fiestas, excursiones… Mis 16. A mis chicas. Irene, mi pequeña hermana y cabra loca, siempre a mi lado, apoyándome incondicionalmente. Yo siempre estaré al tuyo, no lo olvides nunca. Valva, Flavia y Bárbara, gracias por los mil momentos de confesiones, que unidos a una cerveza o vino blanco bien fresquito, sientan mejor. Y por todo el apoyo que me habéis dado durante toda mi vida. Alodia, Cristina, Ángela, Paula, Judith, gracias por preocuparos por mi cada vez que volvía a casa. Llevo tiempo estando lejos, y eso me duele, pero pronto espero tener una súper comida en el huerto, hacer asado y reírnos de los problemas de la vida. A mis chicos, Acha, Maiso, Benito, Balanza, Marcos y Alejandrito. No dejéis de alegrarnos con vuestras tonterías, hacen increíbles todas las quedadas juntos. Pero por favor, meteos un poco menos conmigo, grasias. Fuera de mi grupo de toda la vida estás tú, Pedro. Para mí eres de mi familia, mi hermanito pequeño. No puedo evitar reírme al pensar lo sumamente tocinos que éramos de pequeños, si nos hubieran dejado en una piara a los dos, hubiéramos sido felices. Ha sido muy bonito crecer a tu lado, porque hemos hecho eso, crecer y mucho. Ni se la de consejos que me has dado y me sigues dando, aunque nos hayamos tenido que poner de acuerdo para poder hablar estando tú en América y yo en España. Ya te vale de hacer el pendón por ahí, vuelve a casa amigo, quiero otro verano como el del año pasado a tu lado. Fuera del poblao, y con los años, llegaron otros, muchos otros. Donosín, mi gran amigo del alma, el amigo que todo el mundo desearía tener. Quién iba a decir que aquellas salidas de fiesta en el Cerbuna iban a forjar una amistad así. Han pasado muchos años ya desde que nos conocimos, pero sé que te tendré siempre. Sigue cuidándome como lo haces, nunca dejes de hacerlo. Y cuídame a mi Silvi, una de las personas que más me ha enseñado que a esta vida hay que ponerle una buena sonrisa y echarle narices, por no decir otra cosa. Otro apoyo que seguiré teniendo, a pesar de que a veces pasen meses sin vernos es el de mis Pelillos. Qué bonito es aquello que unió la carrera de Biotecnología, porque la amistad que hicimos estoy segura de que se va a mantener por siempre. Cada vez que nos reunimos es como si el tiempo no hubiera pasado. Riendo, bailando, y poniéndonos al día prestando muchísima atención a lo que cada una de nosotras cuenta que le ha pasado durante esos meses. Cuidándonos. Inés, mi loca preferida, mi gran compañera. Acompáñame con tu risa contagiosa por siempre y dame un poquito de esa flor que llevas contigo. Teresa, gracias por creer en mí como lo haces y hacerme ver en mis momentos de bajón que todo trabajo se recompensa. Rosa, no dejes de venir nunca a Zaragoza, apuntarte a un bombardeo conmigo y alegrarme un fin de semana. María, quiero seguir bailando contigo hasta no poder más y sabiendo que no competimos en las mismas ligas. Carmen, guardo en la cabeza mil y una historias locas que te han pasado. Gracias por contármelas (menos la del garbanzo). Selena, sigue siendo nuestro pequeño desastre que tantas aventuras trae a los Pelillos. Pepelu, el otro día te vi después de 7 años, y guau, qué maravilla amigo. Alicia, qué pequeño pero bonito momento compartimos en la carrera, y tuvimos la suerte de revivir en Marsella. Además, tengo la suerte de tener un grupo de amigos Erasmus, sin jamás haber hecho un Erasmus. Quiero dar gracias a la vida por seguir poniéndome a gente maravillosa en ella de la manera literalmente más aleatoria posible. Darío, Mónica, Riaño, Arturo, María S., Casti, Manu, Ari… Nos quedan muchas celebraciones por vivir, estoy segura. Sara, que bonito lo que unas ganas locas sin coche para ir al CrossFit unió. Gracias por escucharme en tantísimas ocasiones, en tantos paseos y en tantos *¿nos vemos un momento rápido a final del día?* Gracias también a toda la gente que este deporte me permitió conocer, especialmente a mi equipo, Power Schbar, sois muy grandes chicos. Elena, otra maravilla de persona que apareció hace muchos años y que siempre va acompañada de una sonrisa. Kiwis, por supuesto no hubiera terminado haciendo un doctorado si mi experiencia en el Máster de Biomedicina no hubiera sido como fue. Gracias por ser mi hogar en Barcelona. *Non posso nemmeno dimenticare la mia piccola famiglia italiana che mi ha dato tanto amore in un momento molto difficile della mia vita. Siete persone incredibili e presto potrete vedere il mostro in azione.* En este punto digo… madre mía, pero que suerte he tenido. En mi cabeza me siguen viniendo personas que he conocido y con las que sé que compartiré muchos más momentos increíbles. No puedo nombrar a todos, pero estáis ahí. Gracias a aquellos que me sacáis una sonrisa cuando nos reencontramos en el Tony, un festival o en un concierto después de mucho tiempo. Seguiremos bailando juntos.

No todos los agradecimientos van para personas de dos patas. Tengo la suerte de haber podido contar con el mejor secretario durante la gran parte de la escritura de esta Tesis. Sin duda, ese secretario es Melón. Y menos mal que este secretario habla con *meow meow*, no queremos que se desvelen todos los estados de ánimo, momentos de delirio, idas de cabeza y cosas de lo más aleatorias que han podido ocurrir durante la escritura de esta memoria. Ay Meloncito, no te escapes más de casa porfiplis, y sigue siendo mi pequeña gran revolución por muchos años más.

Guillermo, llegué a ti totalmente rota, no me escondo. La vida te va poniendo una mochila, en la que a veces metes demasiadas piedras, y caer abajo no es algo que se elija. Gracias, simplemente gracias, por todo el trabajo el trabajo y dedicación que has tenido conmigo. Creo que realmente no eres consciente de lo que me has ayudado en el final de este periodo, y en muchos otros aspectos de mi vida, para que salieran adelante. Aún nos queda mucho camino por delante y ambos lo sabemos (sorry aún no te vas a librar de mi), pero como tú bien dices, lo importante no son los picos altos de la curva, sino la tendencia positiva con el tiempo. Eso, y que muchos sentimientos son agujetas. Gracias también por hacer que me reencontrase con el poder del deporte, la música y la pintura. Junto a ti, habéis sido mi gran terapia durante mucho tiempo. Me habéis hecho más fuerte.

"*La familia no se elige*". Mítica frase que hemos escuchado hasta la saciedad. Pero es que es así, y puedo decir con orgullo que, si tuviera que elegirse, elegiría una como la mía. Recuerdo con infinito cariño todos los meses de verano que pasaba con mis abuelos Valentina y Marciano en el pueblo. Ay, el pueblo. Qué lloreras y berrinches tenía cuando se acababa esa época y no quería separarme de ellos. Las albóndigas de la abuela, pasar a ver a las gallinas y a los cerditos recién nacidos con el abuelo, y las meriendas en la terraza en las que mil y una vez pedía que me hicieras, abuelo, el plátano-cerdito. Todo eran las aventuras deseadas de un niño. E igualmente doy gracias a mis otros dos abuelos, Flavia y Julián, que me enseñaron lo increíble que podían ser pasar las tardes después del colegio en su huerta con mis amigas. Ay mi abuela, con sus miles de jerseys que me ha hecho de lana con tantísimo cariño y que tan orgullosa llevo, y abuelo, aunque ya no estés con nosotros, doy gracias por haberte podido conocer de niña, pero no lo suficientemente niña, como para no guardar en mi mente pequeños recuerdos de miradas, abrazos y momentos. En uno de mis recuerdos más sólidos te visualizo jugando con tus nietos a la gallinita ciega en el camino de la huerta. Recuerdo perfectamente cómo te reías y la sonrisa que tenías. Prometo que jamás se me olvidará, lo llevo marcado. Gracias al resto de mi familia, a mis primos Darío, Mario, Óscar y Lara, y a mis tíos, Juan, Alodia, Pili, Manuel, Ana y Reyes por quererme tal como soy, aunque haya dado alguna que otra guerra en las comidas familiares. En especial, gracias tía Reyes. De pequeña me criaste como a una hija y me diste todo tu amor, y sigues preocupándote por mí siempre. Si existiera una segunda madre, sin duda eres tú.

Y ahora sí. Quería dejar esto para el final, para sacar lo que realmente está en primer lugar en mi corazón. Quiero dar las gracias a quien les debo todo, no solo todo el apoyo recibido durante esta Tesis, sino todo lo que han hecho, hacen, y sé que harán por mí. Papás, Lucía: esta Tesis es vuestra. Mamá, eres un ejemplo de superación, de constancia total en el trabajo, de implicación con todos tus seres queridos, de apoyo incondicional. Eres un verdadero ejemplo del significado de la palabra *mamá*. Y no te cambio por absolutamente nada. Saber que cada día te tengo a mi lado, a pesar de la distancia, es algo que me tranquiliza mucho y me recuerda día a día donde tengo mi hogar. Creo que no podría contar la cantidad de momentos en los que he estado agobiada, triste, enfadada, disgustada… y siempre sé a quién tengo que recurrir. Y es a ti. Gracias por escucharme a diario, por intentar entenderme siempre, ya sea con aspectos de la vida como por intentar entender los pequeños avances que iba haciendo a lo largo de esta Tesis. Gracias por la frase "*solo un poquito más, cariño*" que me has dicho mil veces y me repito constantemente en mi cabeza. Gracias por preocuparte de mí, por ser *mamá*. Y gracias por cuidar de papá, de la tata y de mí siempre. Papá, si de alguien tengo claro que he sacado mi cabezonería y querer hacer todo perfecto es de ti. Me has enseñado a ser una persona con carácter, extrovertida y concienzuda. Me has enseñado que las cosas no te las regala la vida, que hay que trabajar muy duro para conseguirlas, que muchas veces las cosas no salen como queremos, y que hay que buscar soluciones en vez de llorar y no luchar por ello. Creo sinceramente, que posiblemente seas la persona que más ha creído en mí todos estos años. Jamás has pensado que algo no podía hacerlo. Siempre me has dado ánimo y apoyado en mis decisiones. Me has preguntado por experimentos de mi Tesis una y mil veces, y aunque he tenido que repetirte una y mil veces lo mismo, nunca dejes de preguntarme por los pequeños avances que hago en este mundo. Nunca dejes de confiar en mí y darme todo el apoyo que me das. Yo siempre voy a estar a tu lado, pase lo que pase papá, te lo prometo. Vosotros dos, dos personas increíbles que jamás nos habéis dicho si teníamos que estudiar o no, si era mejor esa carrera o esa salida profesional, si debíamos mejorar nuestro currículum con un máster o un curso. Jamás. Nos habéis dado alas. Pero alas desde la responsabilidad y el aprender a tomar decisiones. Que unos padres te apoyen en tu carrera profesional y personal, hagas lo que hagas, es algo muy, pero que muy grande. Si la tata y yo somos como somos, sin ninguna duda, os lo debemos a vosotros dos. Gracias por ser los padres que sois. Tata, que suerte tenerte pequeña. El primer día que te conocí no quise saber nada de ti, es así, pero desde hace muchos años no puedo dar más gracias a los papas de que me dieran una hermanita como tú. Eres mi hermana sí, pero para mí, si se puede, eres mucho más que eso. Eres mi consejera, mi amiga, la que me dice como son las cosas, la que me apoya, la que me da tranquilidad y calma, e incluso en muchas ocasiones eres mi hermana mayor, la que pone orden y cordura a mi vida, la sensata, la que me ayuda a tomar decisiones, la que me guía. Realmente creo que no eres consciente de todo lo que haces por mí. Eres un pilar clave en mi día a día. Me has ayudado a superar mis épocas difíciles, que a veces no han sido pocas, siempre sacándome una cervecita bien fría y una sonrisa. Ojalá hubiera más personas como tú en el mundo. La vida sería mucho más fácil sin duda. Y además tienes el mérito añadido de haber vivido rodeada durante tanto tiempo de dos personas doctorandas. Gracias por no dejar que Adriano, el mejor cuñi del mundo y yo, nos hayamos rendido. Y con esto sí que sí, ya acabo. De corazón, gracias mamá, papá y tata, por haberme cuidado y guiado cuando ni yo misma creía que podía seguir adelante. Os quiero.

INDEX

INDEX

LIST OF PUBLICATIONS **23**

ABSTRACT **27**

RESUMEN **31**

ABBREVIATIONS **35**

GENERAL INTRODUCTION. Importance of the drug discovery process and the selection of appropriate targets **43**

CHAPTER I. Drug discovery for inhibiting BFT-3: from a chemical screening to an animal model **51**

- *1. INTRODUCTION* *53*
  - 1.1. Colorectal cancer 55
  - 1.2. Genus *Bacteroides* 58
    - 1.2.1. *B. fragilis* 59
      - 1.2.1.1. *B. fragilis*: an overview 59
      - 1.2.1.2. *B. fragilis*: diagnosis 60
      - 1.2.1.3. *B. fragilis* toxin 61
        - 1.2.1.3.1. *B. fragilis* toxin: isoforms 62
        - 1.2.1.3.2. *B. fragilis* toxin: structure 63
        - 1.2.1.3.3. *B. fragilis* toxin: functions and implications 67
      - 1.2.1.4. *B. fragilis* treatment in the context of antimicrobial resistance 71
  - 1.3. Animal models for studying *B. fragilis* infections 74
    - 1.3.1. *G. mellonella* waxworm: an overview of its increasing importance 75
      - 1.3.1.1. General characteristics of *G. mellonella* 76
      - 1.3.1.2. Life cycle of *G. mellonella* 76
      - 1.3.1.3. Reasons for the increased use of *G. mellonella* 77
      - 1.3.1.4. Differences in *G. mellonella* management protocols: different methods for administration 82
      - 1.3.1.5. *G. mellonella* as infection model in pathologies caused by pathogens 82
- *2. OBJECTIVES* *85*
- *3. MATERIALS AND METHODS* *89*
  - 3.1. Plasmid construction 91
  - 3.2. Protein expression and purification 92
  - 3.3. Circular dichroism 94
  - 3.4. Fluorescence spectroscopy 95
  - 3.5. Differential scanning fluorimetry 96
  - 3.6. Differential scanning calorimetry 97
  - 3.7. Chemical libraries 97
  - 3.8. Experimental ligand screening: thermal shift assays 98
  - 3.9. Nomenclature of selected compounds in the *in vitro* experimental ligand screening 99
  - 3.10. Isothermal titration calorimetry 99
  - 3.11. *In vitro* enzymatic inhibition assay 101
  - 3.12. Crystallographic assays 102
  - 3.13. X-ray crystallographic assays: structure determination and analysis 103
  - 3.14. Cell culture assays 104
  - 3.15. Cytotoxicity assays 104
  - 3.16. Microscopy assays 105

- 3.17. Western blot assays 106
- 3.18. *B. fragilis* strains 107
- 3.19. *B. fragilis* cultures 107
- 3.20. Determination of the antibacterial activity of the compounds: liquid minimum inhibitory concentration assays (liquid MIC_{90} assays) 109
- 3.21. Liquid media HTSS (high throughput synergy screening): synergy assays 110
- 2.22. Cell viability assays: MTT assays 111
- 3.23. Time-kill kinetics assays 112
- 3.24. *G. mellonella* waxworm larvae: purchase and maintenance 113
- 3.25. Preparation of *B. fragilis* cultures for *G. mellonella* infections 114
- 3.26. Infection of *G. mellonella* larvae with *B. fragilis* cultures 114
- 3.27. Monitoring of larvae mortality and survival analysis determination 116

4. RESULTS *117*

- 4.1. Structural and functional characterization of BFT-3 and the role of the zinc as cofactor 119
  - 4.1.1. proBFT-3 was correctly purified 119
  - 4.1.2. proBFT-3 is a zinc-dependent conditionally disordered protein 119
- 4.2. Identification of novel bioactive compounds targeting the BFT-3 protein 122
  - 4.2.1. Experimental screening procedure to identify potential ligands targeting the zinc-free partially disordered proBFT-3 state 122
  - 4.2.2. Target engagement of selected compounds by ITC 124
- 4.3. Effect of the selected compounds on the BFT-3 protein using *in vitro* and *in vivo* systems 128
  - 4.3.1. The selected compounds inhibited the enzymatic activity of BFT-3 *in vitro* 128
  - 4.3.2. MOA compounds showed little or no cytotoxic effect on HT-29 and HeLa cells 130
  - 4.3.3. Optical microscopy assays were not determinant in evaluating the effect of MOA compounds 133
  - 4.3.4. Western blot assays revealed that selected compounds inhibit E-cadherin processing in cell cultures *in vitro* 136
  - 4.3.5. Structure determination of inhibitory complexes by crystallographic assays 137
  - 4.3.6. MOA4 showed a greater antimicrobial effect *in vitro*, compared to MOA9 and MOA10 compounds 149
  - 4.3.7. MOA4 may have some synergy effect with MTZ, but not with FOX in *B. fragilis* cultures 152
  - 4.3.8. First steps, establishment of standard conditions and aims in the implementation of *G. mellonella* as a new infection animal model for *B. fragilis* 155
  - 4.3.9. MOA4 was the less toxic compound in *G. mellonella* larvae, followed by MOA9 and MOA10 157
  - 4.3.10. Success in developing a new *G. mellonella* animal infection model to study *B. fragilis* infections 159
  - 4.3.11. The treatment of infected *G. mellonella* larvae with MOA compounds increased their larval survival rate 160

5. DISCUSSION *163*

6. CONCLUSIONS *181*

7. CONCLUSIONES *185*

CHAPTER II. Drug discovery for inhibiting NUPR1: from a chemical screening to an animal model 189

CONTRIBUTIONS *191*

1. INTRODUCTION *195*

- 1.1. Pancreatic cancer 197
- 1.2. Intrinsically disordered proteins 200
  - 1.2.1. NUPR1: an overview 202
    - 1.2.1.1. NUPR1: sequence, isoforms and structure 202

1.2.1.2. NUPR1: functions and implications 203
1.2.1.3. NUPR1-related sequences: NUPR1-like 207
1.2.1.4. NUPR1 and its interaction with biological partners 208
1.3. Molecular mechanism of the classical nuclear protein transport and the role of Importin α 208
1.4. Targeting NUPR1 to treat PDAC: an overview of our research in recent years 211

2. OBJECTIVES *221*

3. MATERIALS AND METHODS *225*
3.1. Plasmid constructions 227
3.2. Protein expressions and purifications 228
3.3. Design, prediction and synthesis of NLS-NUPR1 and NLS-NUPR1L peptides 229
3.4. ITC assays 231
3.5. Chemical libraries 232
3.6. Experimental ligand screening: TSA assays 232
3.7. Nomenclature of selected compounds in the *in vitro* experimental ligand screening 233
3.8. Cell lines and cell cultures 233
3.9. Cell viability assays 234
3.10. hERG channel binding assay 234
3.11. LDH assay, caspase-3/7 activity assay and ATP production assay 235
3.12. Crystal violet 237
3.13. Animals 237
3.14. Statistics 238

4. RESULTS *239*
4.1. Functional characterization of NUPR1 and its interaction with Impα3 as biological partner 241
4.1.1. NUPR1 was correctly purified 241
4.1.2. NUPR1 and Impα3 interacted *in vitro* 241
4.1.3. NUPR1 interacted also with ΔImpα3 *in vitro* 243
4.1.4. The NLS-NUPR1 peptides interacted with both Impα3 and ΔImpα3 *in vitro* 244
4.1.5. NUPR1L interacted with Impα3 and ΔImpα3 *in vitro* 248
4.1.6 The NLS-NUPR1L peptide interacted with both Impα3 and ΔImpα3 *in vitro* 248
4.1.7. ZZW-115 inhibited the nuclear transport of NUPR1 by competing with importins 250
4.2. Identification of novel antineoplastic compounds targeting the NUPR1 protein 252
4.2.1. Experimental screening procedure to identify potential ligands targeting NUPR1 252
4.2.2. Target engagement of selected compounds by ITC 254
4.3. Effects of the selected compounds on the NUPR1 protein using *in vitro* and *in vivo* systems 255
4.3.1. Selected compounds induced a growth suppressive effect in different cell lines 255
4.3.2. AJO14 showed very low hERG binding ability 261
4.3.3. AJO14 induced PC cell death by concomitant necrosis and apoptosis 262
4.3.4. AJO14 treatment induced energetic metabolic failure 263
4.3.5. AJO14 inhibited PDAC tumor development in animal models 264

5. DISCUSSION *267*

6. CONCLUSIONS *287*

7. CONCLUSIONES *291*

REFERENCES **295**

ANNEXES **335**

Annex I: Publications associated to this Doctoral Thesis 337
Annex II: Participation in other projects 399
Annex III: Supplementary tables 403
Annex IV: Supplementary figures 419

LIST OF PUBLICATIONS

This Doctoral Thesis comprises the following scientific publications published, or pending publication. For further details, complete published articles are in the Annex I.

Chapter I is based on two articles:

- **Jimenez-Alesanco A***, Eckhard U*, Asencio del Rio M, Vega S, Guevara T, Velazquez-Campoy A, et al. Repositioning small molecule drugs as allosteric inhibitors of the BFT-3 toxin from enterotoxigenic Bacteroides fragilis. Protein Sci. 2022;31(10):1-18. (*Equal contribution).
- Pending publication: **Jimenez-Alesanco A**, Gomara-Lomero M, Abian O, Ainsa JA, Ramon-Garcia S, Velazquez-Campoy A. Development of Galleria mellonella infection model for the identification of Bacteroides fragilis toxin allosteric inhibitors.

Chapter II is based on three articles and one patent:

- Neira JL, Rizzuti B, **Jimenez-Alesanco A**, Palomino-Schätzlein M, Abian O, Velazquez-Campoy A, et al. A phosphorylation-induced switch in the nuclear localization sequence of the intrinsically disordered nupr1 hampers binding to importin. Biomolecules. 2020;10(9):1-22.
- Neira JL, Rizzuti B, **Jimenez-Alesanco A**, Abian O, Velazquez-Campoy A, Iovanna JL. The paralogue of the intrinsically disordered nuclear protein 1 has a nuclear localization sequence that binds to human importin α3. Int J Mol Sci. 2020;21(19):1-19.
- Pending publication: Liu X*, **Jimenez-Alesanco A***, Zexian L, Neira JL, Peng L, Rizzuti B, Chuluyan E, Velazquez-Campoy A, Xia Y, Abian O, Santofimia-Castaño P, Iovanna J (*Equal contribution). Development of a novel and safe NUPR1 inhibitor with efficient anticancer activity for pancreatic cancer treatment.
- Pending publication patent: Juan L. Iovanna’s group. Centre de Recherche en Cancérologie de Marseille, CRCM. INSERM U1068. Marseille, France. Title: Using NUPR1 as target for identifying small compounds against cancer cell.

Apart from those included in this Doctoral Thesis, additional assays were carried out within the framework of other projects. For this reason, the scientific publications published as a result of these collaborations are listed in the Annex II. In each publication, the research performed by the PhD student is briefly indicated.

ABSTRACT

Drug discovery is a complex process which is focused on developing novel agents that can be used to treat and alleviate diseases that currently lack effective treatments. The identification of drugs against target proteins involved in the development of these diseases is one of the main approaches to combat them. One of the health problems of greatest concern today is cancer.

This Doctoral Thesis includes a combination of several techniques for the identification of potent candidate compounds against two target proteins: BFT-3, associated with colorectal cancer development (Chapter I), and NUPR1, involved in pancreatic cancer development (Chapter II).

Chapter I is related to *Bacteroides fragilis,* a common commensal bacterium in the human colon. However, under dysbiotic conditions, enterotoxigenic *B. fragilis* (ETBF) can emerge, leading to conditions such as diarrhea, anaerobic bacteremia, inflammatory bowel disease, and colorectal cancer. The primary virulence factor of ETBF is a zinc-dependent metallopeptidase called *B. fragilis* toxin (BFT), which damages the intestinal mucosa and initiates disease-related signaling pathways. This Doctoral Thesis was focused on one isoform of BFT, BFT-3. After a structural characterization for evaluating the importance of the zinc for the folded structure of BFT-3, several approved drugs were successfully identified as allosteric inhibitors through a combination of biophysical, biochemical, structural, cellular and *in vivo* animal model techniques. Furthermore, a novel infection animal model in *G. mellonella* larvae for *B. fragilis* infections has been developed, and the results suggest that this new animal infection model could act good host to discover and evaluate new antimicrobials against this pathogen. In summary, these findings show promise for developing treatments for chronic *B. fragilis* infections, diminishing the risk of intestinal inflammation and colorectal cancer development. They can also be further optimized to enhance their affinity, selectivity, or bioavailability. Overall, this approach offers a novel way for developing highly specific inhibitors of ETBF-mediated enteropathogenic conditions.

Chapter II is related to pancreatic cancer, a fatal disease with a high mortality rate with no effective treatment options for patients. A cell stress-related protein, the nuclear protein 1 (NUPR1), has been identified as an essential factor for its development and progression. This protein is an intrinsically disordered protein (IDP) that has a nuclear localization sequence (NLS) to allow for nuclear translocation. This Doctoral Thesis focused on characterizing how NUPR1 translocates to the cell nucleus and studying the role of Importin α3 (Impα3) in its transport from the cytoplasm, by employing several mutants of the NLS region. The NLS region of NUPR1 interacted with Impα3 with high affinity, and this binding was hampered by the removal or phosphorylation of some key residues of NUPR1. Additionally, a similar characterization was performed with a paralogue of NUPR1, the NUPR1L protein, indicating that this other IDP also bound to Impα3 as a mean of being translocated to the cell nucleus. These findings open the way to model the nuclear translocation of disordered proteins. Additionally, since there is an urgent need to identify novel and safe NUPR1 inhibitors, through a combination of biophysical, biochemical, cellular and *in vivo* animal model techniques, a new compound (AJO14) was identified as molecule binding to NUPR1. This compound exerted a potent cytotoxic effect *in vitro* by specifically targeting NUPR1. Mechanistically, it induced cell death mainly through apoptosis, necrosis and promoting a strong metabolism failure with a decrease in ATP production. Furthermore, AJO14 showed an excellent dose-dependent tumor reduction activity on PDAC xenograft mice with no side effects. Overall, AJO14 represents a potent new NUPR1 inhibitor for the treatment of pancreatic cancer.

RESUMEN

El descubrimiento de fármacos es un proceso complejo que se centra en el desarrollo de nuevos compuestos que puedan utilizarse para tratar y aliviar enfermedades que actualmente carecen de tratamientos eficaces. La identificación de fármacos contra proteínas diana implicadas en el desarrollo de estas enfermedades es uno de los principales enfoques para combatirlas. Uno de los problemas sanitarios que más preocupa en la actualidad es el cáncer.

En esta Tesis Doctoral se utilizan diversas técnicas para la identificación de potentes compuestos candidatos contra dos proteínas diana: BFT-3, asociada al desarrollo de cáncer colorrectal (Capítulo I), y NUPR1, implicada en el desarrollo de cáncer de páncreas (Capítulo II).

El Capítulo I está relacionado con *Bacteroides fragilis*, una bacteria comensal en el colon humano. Sin embargo, en condiciones disbióticas, puede propiciarse el crecimiento de *B. fragilis* enterotoxigénicas (ETBF), que provocan afecciones tales como diarrea, bacteriemia anaerobia, enfermedad inflamatoria intestinal y cáncer colorrectal. El principal factor de virulencia de ETBF es una metaloproteasa dependiente de zinc llamada toxina de *B. fragilis* (BFT), que daña la mucosa intestinal e inicia vías de señalización relacionadas con estas enfermedades. Esta Tesis Doctoral se ha centrado en una isoforma de la BFT, la BFT-3. Tras una caracterización estructural para evaluar la importancia del zinc en la estructura plegada de BFT-3, se han identificado con éxito varios fármacos (compuestos MOA), aprobados por la FDA, como inhibidores alostéricos mediante una combinación de técnicas biofísicas, bioquímicas, estructurales, celulares y de modelos animales *in vivo*. Además, se ha desarrollado un nuevo modelo animal de infección de larvas de *G. mellonella* para infecciones por *B. fragilis*, y los resultados sugieren que este nuevo modelo animal podría actuar como un buen huésped para descubrir y evaluar nuevos antimicrobianos contra este patógeno. En resumen, estos hallazgos resultan prometedores para desarrollar tratamientos contra las infecciones crónicas causadas por *B. fragilis*, disminuyendo el riesgo de inflamación intestinal y el desarrollo de cáncer colorrectal. Estos compuestos también pueden optimizarse aún más para mejorar su afinidad, selectividad o biodisponibilidad. En conjunto, este enfoque permite desarrollar inhibidores altamente específicos de las afecciones enteropatógenas mediadas por ETBF.

El capítulo II está relacionado con el cáncer de páncreas, una enfermedad letal con una elevada tasa de mortalidad sin opciones de tratamiento eficaces para los pacientes. Se ha identificado una proteína relacionada con estrés celular, la proteína nuclear 1 (NUPR1), como factor esencial para el desarrollo y progresión de la enfermedad. Esta proteína es una proteína intrínsecamente desordenada (IDP) que posee una secuencia de localización nuclear (NLS) para permitir su translocación al núcleo. Esta Tesis Doctoral se ha centrado en caracterizar cómo NUPR1 se

transloca al núcleo celular y estudiar el papel de la Importina α3 (Impα3) en su transporte desde el citoplasma, empleando varios mutantes de la región NLS. Se ha observado que la región NLS de NUPR1 interacciona con Impα3 con gran afinidad, y esta unión se vio dificultada por la eliminación o fosforilación de algunos residuos clave de NUPR1. Además, se ha realizado una caracterización similar con un parálogo de NUPR1, la proteína NUPR1L, indicando que esta otra IDP también se une a Impα3 para ser translocada al núcleo celular. Estos hallazgos abren el camino para modelizar la translocación nuclear de proteínas desordenadas. Además, dado que existe una necesidad urgente de identificar inhibidores de NUPR1 novedosos y seguros, mediante una combinación de técnicas biofísicas, bioquímicas, celulares y de modelos animales *in vivo*, se ha identificado un nuevo compuesto (AJO14) que interacciona con NUPR1. Este compuesto tiene un potente efecto citotóxico *in vitro* al dirigirse específicamente a NUPR1. Respecto a los mecanismos moleculares de acción, induce muerte celular principalmente a través de procesos de apoptosis, necrosis y promueve un fuerte fallo del metabolismo asociado a una disminución de la producción de ATP. Además, AJO14 reduce el crecimiento tumoral de forma dosis dependiente en ratones xenoinjertados con PDAC sin mostrar efectos secundarios. En general, AJO14 representa un nuevo y potente inhibidor de NUPR1 para el tratamiento del cáncer de páncreas.

ABBREVIATIONS

ΔC_P Unfolding heat capacity

ΔG Gibbs energy variation

ΔH Enthalpy variation

$\Delta H(T_m)$ Unfolding enthalpy variation

ΔImpα3 Truncated importin α3 without the Importin β-binding domain

ΔS Entropy variation

ε Molar extinction coefficient

λ_{exc} Excitation wavelength

λ_{em} Emission wavelength

5’-UTR 5’-untranslated region

Å Angstrom unit

A375 Human melanoma cells

a.u. Absorbance units

ac Acinar cells

Abs Absorbance

ACT Acetate

ADMET Chemical absorption, distribution, metabolism, excretion and toxicity

Amp Ampicillin

AMPs Antimicrobials peptides

ANOVA Analysis of variance

ANS 8-anilino-1-naphthalenesulfonic acid

ARM Armadillo

ATF-4 Activating transcription factor 4

ATM Ataxia-telangiectasia mutated gene

AURKA Aurora kinase A

B. fragilis Bacteroides fragilis

Bcl-2 B-cell lymphoma 2 protein

bft Bacteroides fragilis toxin gene

BFT *Bacteroides fragilis* toxin

BFT-3 *Bacteroides fragilis* toxin isoform 3

BRCA2 Breast Cancer Type 2 gene

BRCA1 Breast Cancer Type 1 gene

BSA Bovine serum albumin

C-RING1B C-terminal domain of RING1B

cal calorie

CatD Catalytic domain

CBA Checkerboard assays

CD Circular dichroism

CDKN2A Cyclin-dependent kinase inhibitor 2A gene

CDP Conditionally disordered protein

CEC Colonic epithelial cell

CFU Colony-forming unit

CHOP C/EBP homologous protein

Com-1 Candidate of metastasis 1 (another name for NUPR1)

CRC Colorectal cancer

CTn Conjugative transposon

Cyt*c* Cytochrome C

Da Dalton

DMEM Dulbecco's modified eagle's medium

DMSO Dimethyl sulfoxide

Dnmt1 DNA (cytosine-5)-methyltransferase 1

DSC Differential scanning calorimetry

DSF Differential scanning fluorimetry

du Duct cells

E. coli Escherichia coli

EDO Ethylene glycol

EDTA Ethylenediaminetetraacetic acid

ERS Endoplasmic reticulum stress

ETBF Enterotoxigenic *Bacteroides fragilis*

FBS Fetal bovine serum

FDA Food and Drug Administration

FICI Fractional inhibitory concentration index

FRA Fragilysin (another name for BFT)

FOBT Fecal occult blood tests

FOR Formate anion

FOX Cefoxitin

FPLC Fast protein liquid chromatography

G. mellonella Galleria mellonella

H&E Hematoxylin and eosin stain

H358 Human lung carcinoma cells

HAT Histone acetyltransferase

HCl Hydrochloric acid

HCC Hepatocellular carcinoma

HeLa Henrietta Lack´s cells

HepG2 Human hepatocellular carcinoma cells

hERG Human ether-à-go-go-related gene

HPC High performance computing

HT-29 Human colon adenocarcinoma cells

HTS High throughput screening

IBB Importin β-binding

IC_{50} Half maximal inhibitory concentration

IDPs Intrinsically disordered proteins

IEC Intestinal epithelial cell

IMAC Immobilized metal affinity chromatography

Impα Importin α

Impα1 Importin α1

Impα3 Importin α3

Impα4 Importin α4

Impα5 Importin α5

Impα6 Importin α6

Impα7 Importin α7

Impα8 Importin α8

Impβ Importin β

i.p. Intraperitoneally

IPTG Isopropyl β-D-1-thiogalactopyranoside

ITC Isothermal titration calorimetry

K_a Association constant

Kan Kanamycin

K_d Dissociation constant

kDa Kilodalton

KO Knockout

KRAS Kirsten rat sarcoma virus gene

LB Luria-Bertani broth

LDH Lactate dehydrogenase

LPS Lipopolysaccharide

LTA Lipoteichoic acid

MAPK Mitogen-activated protein kinases

MD Molecular dynamics

MDA-MB-231 Human breast adenocarcinoma cells

MiaPaCa-2 Human pancreatic cancer epithelial cells

MIC Minimal inhibitory concentration

MIC$_{90}$ Minimal inhibitory concentration 90

mP Membrane polarization

MPs Metalloproteases

MTZ Metronidazole

MW Molecular weight

n Stoichiometry for an interaction

NaCl Sodium chloride

NDA New drug application

NEAAs Non-essential amino acids

NF-κB Nuclear factor kappa B

NLS Nuclear localization sequence

NMR Nuclear magnetic resonance

NPC Nuclear pore complex

NTBF Non-enterotoxigenic *Bacteroides fragilis*

nupr1 nuclear protein 1 gene

nupr1L nuclear protein 1-like gene

NUPR1 Nuclear protein 1

NUPR1-KO Knockout *nupr1* Panc-1 clones

NUPR1-WT Wild-type *nupr1* Panc-1 clones

NUPR1L Nuclear protein 1-like

NUPR2 Nuclear protein 1 (another name for NUPR1L)

OD$_{600}$ Optical density at 600 nanometers

ORTH Orthorhombic space group

p8 Molecular mass 8 kDa (another name for NUPR1)

P53BP1 p53 binding protein

PAIs Pathogenicity islands

Panc-1 Human pancreatic cancer cell line

PanINs Pancreatic intraepithelial neoplasms

PC Primary compound

PC-3 Human prostate cancer cells

PCR Polymerase chain reaction

PD Prodomain

PDAC Pancreatic ductal adenocarcinoma

PDB Protein data bank

PEG Diethylene glycol

PGE Triethylene glycol

pI Isoelectric point

pKa Acid dissociation constant

PLA Proximity ligation assay

pLDDT Per-residue confidence score

PMSF Phenylmethylsulfonyl fluoride

PRO L-proline

ProTα Prothymosin α

PTMs Post-translational modifications

qPCR Quantitative real-time PCR

RANKL Receptor activator of NF-κB ligand

RIPA Radioimmunoprecipitation assay buffer

rpm Revolutions per minute

ROS Reactive oxygen species

RSRZ Real-space R-value Z-score

RT-PCR Real-time polymerase chain reaction

SAR Structure-activity relationship

s.c. Subcutaneously

SC Secondary compound

SD Standard deviation

SDS Sodium dodecyl sulphate

SDS-PAGE Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC Size-exclusion chromatography

SMAD Small mothers against decapentaplegic

SMAD4 SMAD family member 4 gene

SSS Sterile saline solution

TBS Tris-buffered saline

TBS-T Tris-buffered saline with Tween 20

TETR Tetragonal space group

TGFβ Transforming growth factor β

THC D9-tetrahydrocannabinol

T_m Midpoint temperature/Unfolding temperature

TFP Trifluoperazine

TK Tyrosine kinases

TP53 Tumor protein P53 gene

TRB3 Telomere repeat-binding factor 3

Tris Tris(hydroxymethyl)aminomethane

TSA Thermal shift assay

U2OS Human bone osteosarcoma epithelial cells

U87 Human primary glioblastoma cells

UV Ultraviolet

WT Wild-type

GENERAL INTRODUCTION. Importance of the drug discovery process and the selection of appropriate targets

Drug discovery is aimed at identifying and developing new drugs to both treat and palliate currently untreated diseases that affect the world's population. Drug discovery is a complex process that nowadays involves many different phases, processes, and agents, and it began in ancient times by using natural products, mainly extracts obtained from plants, for medical purposes. The development and further refinement to substances of these compounds obtained from plant sources has had a great impact on medicinal chemistry. In modern times, the first notable period of drug discovery was in the 19th century when the basis of drug identification relied on the serendipity of medicinal chemists. Subsequently, the second period began in the early 20th century when new drug structures were found that contributed to the new era of antibiotic discovery. Thanks to the knowledge of structures and the development of powerful new techniques, many advances in drug discovery were made during the end of that century. This aspect, together with the emergence of recombinant DNA technology and the Omics revolution in the 21st century, led to the establishment of the third period of drug discovery in which there is a great increase in biopharmaceutical drugs approved by the FDA for therapeutic use (1).

Regarding the different stages of the drug discovery process (**Figure 1**), the first stage is the early drug discovery, in which compounds are identified and tested first *in vitro* and then *in vivo* in animal models to evaluate toxic and pharmacological effects, as well as genotoxic effects. Thus, a set of compounds is proved to be active against a target and a disease-relevant experimental model based on cells, tissues or animals. Research at this point is performed in the laboratory using *in silico* platforms, biochemical assays, cell cultures and animal models. In the preclinical phase, the substances identified during early drug discovery are refined, optimized, and extensively tested in laboratory and in animal or alternative models. If a given compound is considered promising, it is mandatory to present the results obtained from short-term toxicity testing in at least two animal models. Then, Phase I clinical trials are conducted on a small number of healthy volunteers to determine the safe and non-toxic dose range of the compound. If after this phase, the compound is still considered promising, it advances to Phase II, in which the number of patients, who are still volunteers but subject to the medical condition the product is intended to treat, is increased. If the compound continues to show promise after Phase II, it proceeds to Phase III, in which the number of subjects, who present the disease of interest, is much larger. The goal is to test different doses, demonstrate the efficacy of the compound, and establish the optimal compound dose. In addition, potential interaction effects with other medications will also be assessed. Because the number of subjects is larger than that in Phase II trials, it is very likely to observe possible adverse effects produced by the compound. Before a drug can be sold, approval from a national regulatory authority or centralized process is required. If at the end of Phase III the compound is still promising, the manufacturer will file a New Drug Application (NDA) with the Food and Drug Administration (FDA). Finally, Phase IV assays, also known as post-marketing surveillance trials, take place after receiving marketing authorization from the authorities. In this stage, more comprehensive data can be gathered regarding the effectiveness and safety of the new drug because the much larger number of patients taking the drug provides lots of data, as well as the comparison with already available treatments. These studies are designed to assess the long-term effects of a drug in order to keep records as well as to avoid adverse events (2).

From the above, it is indisputable that drug development is a tedious and a resource- and time-consuming process. Furthermore, it also involves a great deal of bureaucratic red tape and exhaustive monitoring to ensure both safety and efficacy, all of which are necessary for the FDA approval of a new chemical entity.

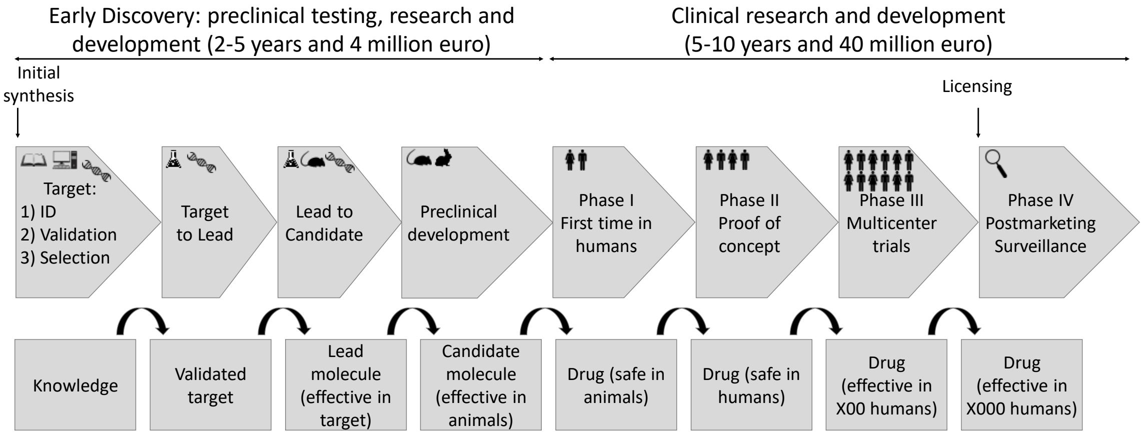

Early Discovery: preclinical testing, research and development (2-5 years and 4 million euro)
Clinical research and development (5-10 years and 40 million euro)
Initial synthesis
Licensing
Target:
1) ID
2) Validation
3) Selection
Target to Lead
Lead to Candidate
Preclinical development
Phase I First time in humans
Phase II Proof of concept
Phase III Multicenter trials
Phase IV Postmarketing Surveillance
Knowledge
Validated target
Lead molecule (effective in target)
Candidate molecule (effective in animals)
Drug (safe in animals)
Drug (safe in humans)
Drug (effective in X00 humans)
Drug (effective in X000 humans)

Figure 1. The drug discovery process.

From the initial stages of drug discovery, until a single drug is available for purchase on the market, it is estimated that 12 years are needed with a budget that could excess 1 billion euros on average. For this reason, identifying and validating drug targets is an initial but extremely important step in drug discovery; starting with good candidate compounds guarantees the accurate direction in subsequent screenings and developments, thus optimizing the identification of more effective and safer drug candidates (2,3). Early rejection of inappropriate candidate compounds is key to reduce drug attrition and failure, and avoid cumbersome and costly procedures towards a dead end route. Besides, typically only 1 out of every 5,000 drugs makes it to the market approval stage and moreover, out of 5,000 and 10,000 drug candidates, only 250 (5-10%) make it to preclinical testing (**Figure 2**).

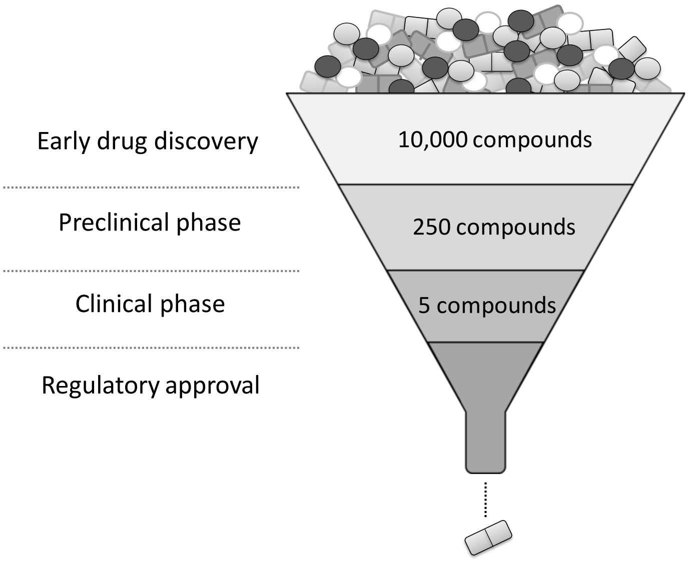


Figure 2. Number of compounds in each stage of the drug discovery process.

For a long time, drug discovery has been focused on the functional modulation of druggable proteins related to malfunctioning protein signaling networks involved in diseases or pathologies. Thus, bioactive small molecules have been used to control or compensate altered protein signaling networks and bring about a reversal from a disease state to a healthy state. That is why the vast majority of the targets of approved drugs are proteins (4). Complex protein networks are essential for sustaining cellular systems due to the fact that many proteins are associated with key cellular processes: signaling, transcription, translation, homeostasis, and cell cycle. Under certain conditions, there are environmental perturbations in cells, which can properly respond to these changes, retaining the organisms their correct physiology, but only when these protein signaling networks are properly modulated and regulated. The dysregulation of these protein signaling cascades favors the transition from a healthy state to a disease state, consequently leading the cellular system to threaten the whole organism (5).

Given the important role of proteins in many cellular processes, and their enormous utility in drug discovery, it is essential to correctly select the target proteins to be used in a drug screening project, in order to ensure that drugs are sought against proteins that are really important for a determined pathology. In addition, given the great complexity, duration and enormous costs required by the drug discovery process, it is essential to establish a good strategy based on the techniques to be used in the first steps. Therefore, it is crucial to know which techniques and equipment are most beneficial and useful in the study to identify compounds against a target protein.

In relation to the above, there is no doubt that one of the health problems of greatest concern nowadays, both to the general population and to researchers, is cancer. Cancer is a puzzling and frightening disease (or, more precisely, a group of diseases) with a great impact and socio-economic burden on citizens, and one of the main causes of death. According to the World Health Organization (WHO), it is estimated that there are more than 19 million new cancer cases every year worldwide (6), expected to increase to 24 million cases every year by 2035.

Breast and lung cancers are the most common cancers worldwide, and colorectal cancer (CRC) is the 3rd most common cancer with more than 1.9 million new cases in 2020, although there has been an increase in five-year survival rates due to diagnostic and therapeutic advances (7,8). In terms of mortality, CRC is also one of the most lethal cancers. Meanwhile, regarding the mortality, although pancreatic cancer (PC) is not among the cancers with the highest incidence, its silent development and rapid aggressiveness makes its mortality rate disproportionately high (9) (**Figure 3**).

Therefore, taking into account this information, two of the most relevant gastrointestinal cancers at present, due to their socio-economic impact, are CRC and PC (10–13). In general, the pathophysiology of gastrointestinal disorders is very complex, involving dysregulation of brain-gut interactions, microbial dysbiosis at the gut level, visceral hypersensitivity, altered mucosal immune functions, as well as alterations in gastrointestinal motility (14,15), and there are a number of symptoms attributed to these gastrointestinal tract problems such as abdominal pain, diarrhea, nausea, vomiting, constipation, bloating, a feeling of fullness, decreased appetite, indigestion, and changes in bowel habits, among others (12,13).

Consequently, in addition to the problems that these disorders pose for the affected population, causing a greatly impaired quality of life, and the great burden of these disorders, the direct and indirect economic impact is enormous (billions of euros annually) (15). Besides, gastrointestinal cancer patients, and certainly overall general cancer patients, have increasingly become chronic patients, with the corresponding rise in health care costs (16).

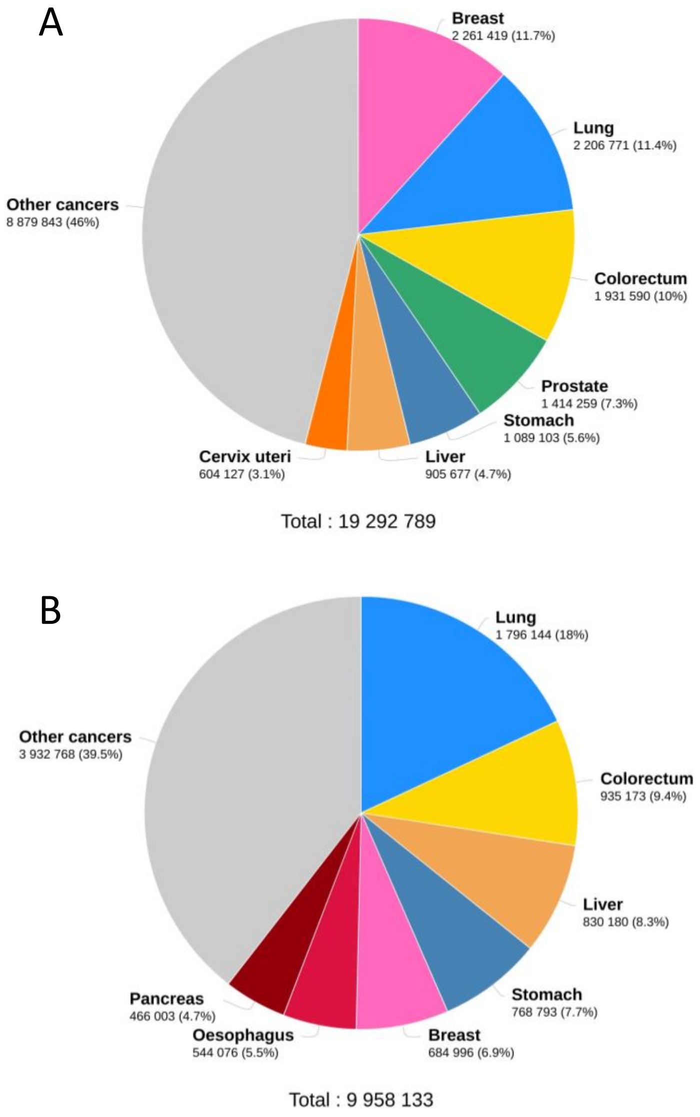


Figure 3. Estimated number of new cancer cases in 2020 (world, both sexes, all ages). (A) Incidence and (B) mortality of cancer worldwide. Figure from: Globocan 2020. Graph production: Global Cancer Observatory. Institutional Agency for Research on Cancer (World Health Organization).

As the world's population continues to grow and age, the global number of cancer deaths continues to rise. With this growing global burden, cancer (and cancer-related diseases, such as other gastrointestinal disorders) prevention and management, are some of the most important public health challenges of the 21st century. Therefore, due to the great social, economic and health importance of cancer, it is urgent to develop new, more effective and specific treatments, improving both quality of life and life expectancy in patients. In this general context, developing drugs against specific proteins involved in gastrointestinal cancers, whether overexpressed, underexpressed or mutated, is one of the ways forward to combat these deadly diseases that affect so many people worldwide.

CHAPTER I. Drug discovery for inhibiting BFT-3: from a chemical screening to an animal model

1. INTRODUCTION

1.1. Colorectal cancer

One of the most important gastrointestinal diseases is CRC. It is the 3rd most common cancer diagnosis and the 2nd deadliest malignancy for both sexes combined, affecting approximately 135,439 estimated new patients in United States per year. In Spain, CRC is the tumor with the highest incidence, affecting 25,000 new cases every year (17). These rates depend on the geography with the highest rates in the most developed countries (**Figure 4**). Because there is a steady progress in developing countries, the incidence of CRC is predicted to increase to 2.5 million new cases worldwide in 2035 (9,18).

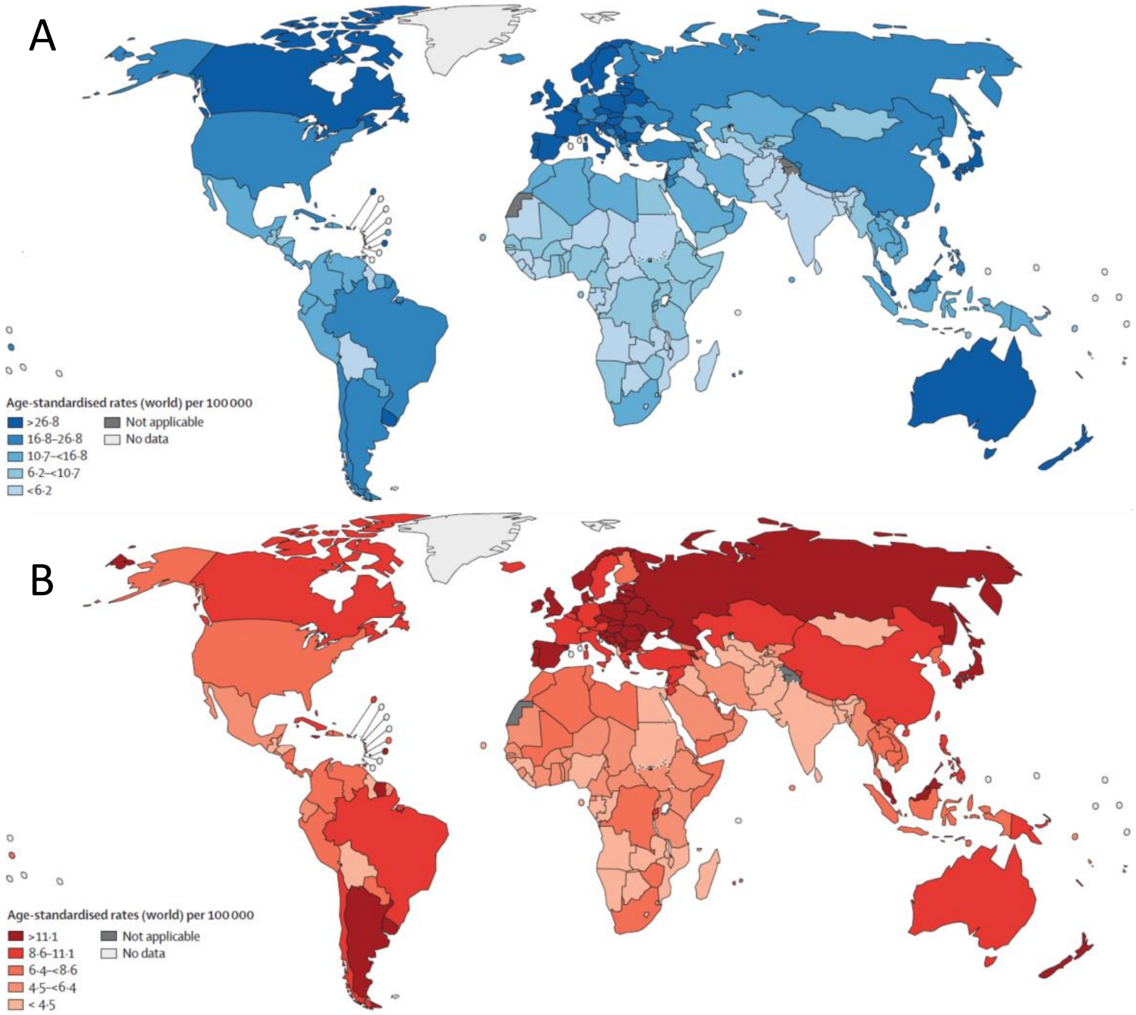

A
Age-standardised rates (world) per 100 000
>26·8
16·8–26·8
10·7–<16·8
6·2–<10·7
<6·2
Not applicable
No data
B
Age-standardised rates (world) per 100 000
>11·1
8·6–11·1
6·4–<8·6
4·5–<6·4
< 4·5
Not applicable
No data

Figure 4. Age-standardized CRC (A) incidence and (B) mortality rates for countries in five continents according to the latest WHO International Agency for Research on Cancer figures. Figure adapted from (18).

Based on this data, CRC screening for average and high-risk populations has been recommended by the European Council. Among the adequate strategies considered for CRC screening, fecal occult blood tests (FOBT) every year, flexible recto sigmoidoscopy every 5 years, and colonoscopy every 10 years are the most relevant for the average-risk population (17,19,20). Thanks to these strategies, the CRC incidence rate has been declining 3% per year since 2004, although increasing by 2% per year among screened young adults (< 50 years) (16).

This type of cancer has both genetic risk factors as well as strong environmental associations. In particular, it could be presented as sporadic cancer (70% of the cases), familial clustering (20%) and inherited syndromes (10%). In most cases, there are no symptoms of CRC in its early stages, but, basically, CRC is initiated when the cells in the colon grow, typically from a small non-cancerous growth on the colon wall (polyp), and multiply uncontrollably (the polyp grows larger and become cancerous), producing damage in the surrounding tissues and thus interfering with the normal functions of the colon (20). Besides, as polyps grow, they can bleed and/or obstruct the intestine. The transformation of the normal colonic epithelium to a precancerous lesion, leading to an invasive carcinoma requires an accumulation of genetic mutations either in the somatic line (acquired) and/or germline line (inherited) within approximately 10-15 years (**Figure 5**). For this reason, the early detection- or removal of polyps would help in reducing the incidence of CRC (19,21).

Regarding to the symptoms and signs of CRC, the most important are: rectal bleeding is one of the most common symptoms, generally regarded as an early CRC symptom; changes in bowel habits arising as changes in frequency of defecation, consistency and shape of the stool and difficulty in evacuation; and unintentional weight loss and abdominal pain. Among the signs, rectal and abdominal palpable masses, iron deficiency anemia, acute symptoms, and metastatic disease are the most important (22).

In terms of treatment, there are several options to try to combat this disease. These include proper endoscopic follow-up, surgical local excision, systemic therapy and downstaging preoperative radiotherapy, extensive surgery for both locoregional and metastatic disease, local ablative therapies for metastases, and immunotherapy, palliative chemotherapy, and targeted therapy. Some of the FDA-approved drugs for the treatment of metastatic CRC are: leucovorin, fluorouracil, irinotecan, capecitabine, oxaliplatin, cetuximab, bevacizumab, panitumumab, trifluridine, tipiracil, pembrolizumab, nivolumab, ipilimumab and encorafenib. Unfortunately, metastatic CRC remains incurable for most patients and, based on several studies from the National Cancer Institute, 14% is the 5-year survival rate for metastatic CRC (18,23,24).

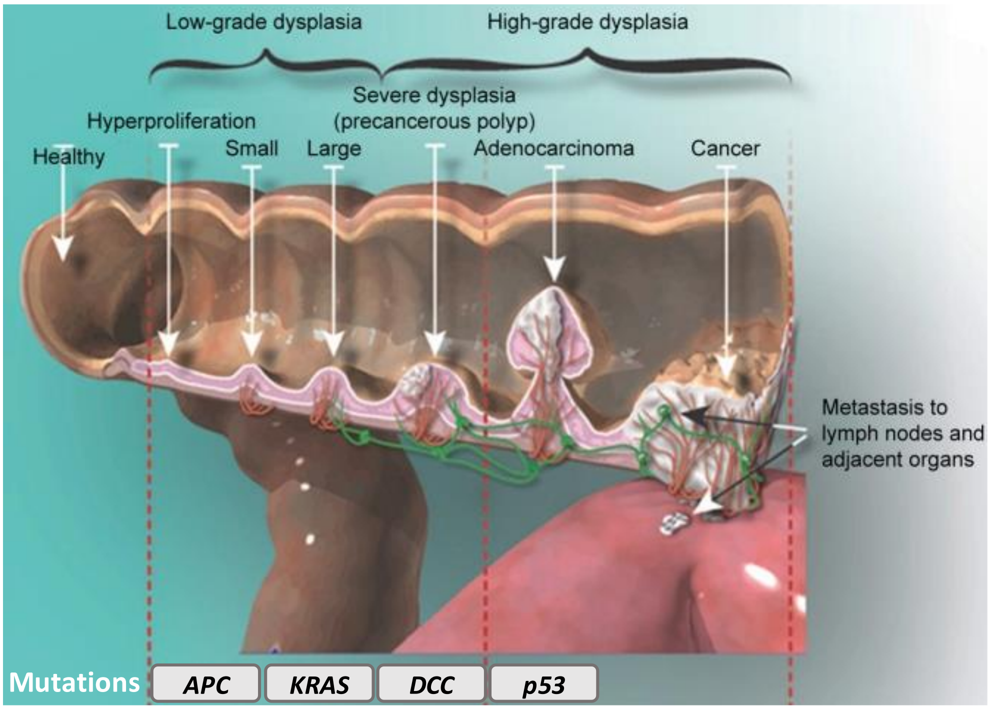

Low-grade dysplasia
High-grade dysplasia
Severe dysplasia
(precancerous polyp)
Hyperproliferation
Healthy
Small
Large
Adenocarcinoma
Cancer
Metastasis to
lymph nodes and
adjacent organs
Mutations
APC
KRAS
DCC
p53

Figure 5. CRC stages and development. In the image the histological changes are shown. There are four stages in the development CRC carcinogenesis: initiation, promotion, progression and metastasis. 1) Initiation (30-60 years): from normal epithelium to small benign growth (polyp). Loss of tumor suppressor *APC* gene. 2) Promotion (10-20 years): from small benign growth (polyp) to large benign growth (early adenoma). Activation of *KRAS* oncogene. 3) Progression (10-20 years): from large benign growth (early adenoma) to large benign growth (late adenoma). Loss of tumor suppressor *DCC* gene. 4) Metastasis (0-5 years): from large benign growth (late adenoma) to malignant tumor (carcinoma). Loss of tumor suppressor *p53* gene. Additional mutations not shown also lead to CRC (20,25). Figure adapted from (25).

Higher CRC risk is associated with both modifiable as well as non-modifiable factors. Taking into account the first type, obesity, smoking, bacterial infections, alcohol intake and consumption of processed and red meat seem to be the most important. On the other hand, physical activity, non-steroidal anti-inflammatory drugs, postmenopausal hormone therapy, and vegetable and fruit intake are associated with a decreased risk of CRC. Among the non-modifiable ones, age, a family history of CRC and inflammatory bowel disease (IBD) are the most relevant. Furthermore, faulty genes may be one of the most important causes of CRC, and it has been estimated that defective genes cause at least 30% of CRC. There are other origins that were initially less clear, but that have gradually gained relevance in studies on the etiology of cancer (26).

As previously mentioned, several infectious agents are considered to be one of the causes for some cancers in humans. Since the 1980s, the interest in investigating the relationship between infections and cancer has increased. In 2002, it was stated that infectious agents are implicated in approximately 18% of all cancers worldwide. This estimation was based on the burden of diseases associated with cancers which have known infectious etiologies, such as liver, cervical and gastric cancers (27).

The human intestine provides a habitat rich in nutrients, allowing the growth of many species of bacteria. Concretely, the highest concentration of bacteria is found in the colon (28), which are normally commensal microorganisms having a symbiotic relationship with their host and important metabolic activities. Among other important roles, the following stand out: recovery of energy and absorbable nutrients, protection of the colonized host against external microbes/pathogens, and important trophic effects on intestinal epithelia and on immune structure and function. However, under certain situations, some microbes that are normally or incidentally found in the colon become pathogenic if they breech the host mucosal barrier translocating into the bloodstream (28,29).

There are 300-500 different species of bacteria in the intestinal habitat of an individual adult, of which 30-40 species comprise up to 99% of the total bacterial population. The dominant genera are *Bacteroides*, *Bifidobacterium*, *Eubacterium*, *Clostridium*, *Lactobacillus*, *Fusobacterium* and various anaerobic Gram-positive cocci (29,30).

At the beginning of the malignant process, when an infiltration by gut bacteria of the normally sterile peritoneal cavity occurs, aerobes such *Escherichia coli* dominate the infection. However, once sufficient oxygen has been depleted, *Bacteroides* species typically dominate and dictate the chronic stage (31).

1.2. Genus *Bacteroides*

As mentioned above, the largest population of bacteria in the human body is present in the colon, and the majority of these organisms are anaerobes. Of these, approximately the 25% are species of the genus *Bacteroides*. *Bacteroides* species are anaerobic, bile-resistant, non-spore-forming, Gram-negative rods (31).

Bacteroides species may be acquired by child from mother during vaginal birth, and from that moment, these bacteria become part of the human flora, being commensals, in the earliest stages of life (32). However, under certain conditions, such as the rupture of the gastrointestinal tract or intestinal surgery, *Bacteroides* organisms can escape from their proper localization, causing significant pathologies, including abscess formation in multiple body sites, such as the abdomen, brain, liver, pelvis, and lungs, as well as bacteremia. Under these situations, these organisms are responsible for infections with significant morbidity and mortality (31,33,34).

Although *Bacteroides fragilis* constitutes only 0.5-2% of the normal colonic flora, it is the most frequent isolate from clinical specimens and is considered the most virulent among *Bacteroides* species (35). Therefore, species of this genus are important clinical pathogens, having an associated mortality rate of approximately 19%, and, importantly, untreated *B. fragilis* infections are associated with mortality rates of about 60% (31,36–38).

1.2.1. *B. fragilis*

1.2.1.1. *B. fragilis*: an overview

Because *B. fragilis* belongs to the genus *Bacteroides*, *B. fragilis* are anaerobic bacilli, Gram-negative, pleomorphic to rod-shaped bacteria. Under appropriate conditions, which will be discussed later, they have good to excellent growth, forming grey colonies surrounded by small brown to black areas (**Figure 6**). *B. fragilis* is usually a commensal organism, meaning that they are part of the normal human colon flora, playing a complex role in the colon and having a beneficial relationship with the host. Concretely, *B. fragilis* strains are present approximately in 0.5-2% of the normal colonic flora and, in up to 80% of both children and adults, strains of this bacterium are identified in the colonic flora (31,39). However, if there is a breach in the integrity of the mucosal lining, making the mucosal barrier disrupted, it can result in abscess formation and bacteremia, among other problems, producing a significant morbidity. Recent surgery, trauma and malignancy are predisposing factors for *B. fragilis* infections. Depending on the different tissues involved, different clinical presentations exist. These bacteria have been isolated from chronic pressure ulcers and diabetic foot ulcers, are involved in inflammatory diarrhea and may be present in necrotizing fasciitis, synergistic anaerobic infections and bloodstream infections, among others (31).

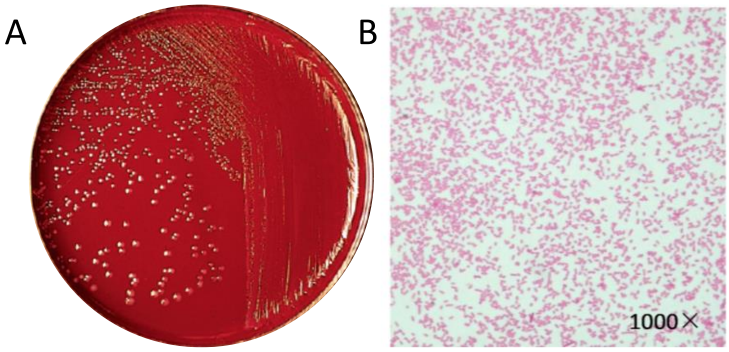


Figure 6. Morphological characteristics of *B. fragilis*. (A) Colonies on *Brucella* agar supplemented with 5% defibrinated sheep blood, 5 µg/mL hemin and 1 µg/mL vitamin K1 in anaerobic conditions at 37 °C. (B) *B. fragilis* bacteria observed under an optical microscope after Gram-negative staining. Figure adapted from (40).

One aspect that partially explains the successful colonization of the mucosa by *B. fragilis,* is that it is relatively aerotolerant. This means that it can grow in the presence of nanomolar concentrations of oxygen compared to other strict anaerobes. Thus, it is able to survive in the mucosa, where oxygen partial pressure is higher, as well as to induce bacteremia (12,41).

B. fragilis strains eliciting intestinal secretion were called enterotoxigenic *B. fragilis* (ETBF), while their non-secretory counterparts were named nontoxigenic *B. fragilis* (NTBF) (42). The main differences between ETBF and NTBF are the presence of *B. fragilis* toxin (BFT), encoded by the *bft* gen, and the ability to produce biofilm (43,44). Then, *B. fragilis* virulence is mostly induced by the BFT toxin production, which is secreted, mediating in this way intestinal inflammation. BFT is yet considered the only known specific virulence factor of ETBF (44). Concretely, the association between ETBF and diarrheal disease was first noted in livestock and then reported in human diarrheal disease in 1987 (45–47).

As mentioned above, microbial imbalances may lead to the activation of chronic inflammation, production of toxins causing DNA damage, and alterations in the gastrointestinal microenvironment. Then, microbes are able to promote cancer development through several routes under certain situations (27,48).

When a chronic ETBF colonization occurs in the intestine, it stimulates chronic intestinal inflammation, and, then, modulating signaling pathways are turned on, such as signal transducers and activators of transcription 3 (STAT3) activation, which contributes to interleukin (IL)-17 production that is involved in colon inflammation (47). Apart from altering signaling pathways, BFT can lead to the production of reactive oxygen species (ROS) that produce DNA damage and it is also responsible for the cleavage of E-cadherin (43,44,49). Therefore, ETBF is an important source of chronic inflammation, is epidemiologically associated with acute diarrheal diseases in humans and livestock, inflammatory bowel disease (IBD), and is considered as a risk factor for CRC (43,44,50,51). Furthermore, the *bft* gene has been detected from colorectal neoplasia in mucosal samples of patients, compared with individuals undergoing outpatient colonoscopy (controls) (52). For instance, some studies have demonstrated that ETBF infection alone is necessary and sufficient to induce acute and persistent colitis in wild-type (WT) C57BL/6 mice (53,54).

1.2.1.2. *B. fragilis*: diagnosis

Regarding the diagnosis of ETBF infections, simple methods do not exist, and a combination of bacterial isolation and the detection of the *bft* gene, or *in vitro* evaluation of the BFT biological activity, are the most commonly used techniques today (**Table S1, Annex III**). However, it is worth noting that, from stool samples, the diagnosis can be challenging (55). This is mainly due to the fact that it requires an anaerobic stool culture, which means to culture *B. fragilis* in anaerobic conditions, representing a complication related to the delays in the process of stool sample processing in hospitals. Besides, working with anaerobic microbes is generally difficult and the presence of a huge heterogeneity of *B. fragilis* strains in fecal are aspects that contribute to the complexity of diagnosis. Another fact is that both NTBF and ETBF can be found in stool samples. Therefore, in order to ensure an accurate diagnosis, it is recommended to employ PCR to detect the presence of the *bft* gene associated with a ETBF infection (42,56,57). PCR based on specific genes of *B. fragilis*, such as *leu*B (β-isopropylmalate dehydrogenase) and *gyr*B (B-subunit of DNA gyrase), apart from *bft* gene, are mostly employed (58,59). Recently, it is also employed real-time PCR (RT-PCR) for amplifying 16S rRNA regions conserved among *B. fragilis* group species (60).

Improving and developing detection methods for the ETBF diagnosis in clinical laboratories is essential. This is why nowadays there are considerable efforts to simplify the techniques that allow an accurate diagnosis. Some of the methods under development are:

- TaqMan RT-PCR based on high-throughput direct detection of ETBF from stool and other clinical samples.
- Digital PCR (dPCR) technology for a better quantification of ETBF directly in the gut microbiome.
- Culturomics, which is a new culture based system that combines various antibiotics or inhibitors, for obtaining a selective growth of bacterial species present in low quantities in the gut population, with matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) or rRNA sequencing for identification (61).
- Enzyme-linked immunosorbent assays (ELISA) to detect ETBF by measuring higher levels of both IgA and IgG against BFT in patient samples compared to controls (62).

1.2.1.3. *B. fragilis* toxin

As mentioned above, ETBF is characterized by a single unique virulence determinant, the production and secretion of *B. fragilis* toxin (BFT; also called fragilysin (FRA)) (35,63).

BFT is a preproprotein metalloprotease enterotoxin (proBFT) with a molecular weight (MW) of 44.4 kDa, encoded by the chromosomal *bft* gene, that is processed and secreted into the culture supernatant by ETBF as the mature 20-kDa protein BFT (31,50).

1.2.1.3.1. *B. fragilis* toxin: isoforms

There are three known BFT isoforms, encoded by the genes *bft1*, *bft2* and *bft3*, being called BFT-1, BFT-2 and BFT-3 proteins, which have pairwise sequence identities of 93-96% (**Table 1**). The immature isoforms are also called proBFT or profragilysins (64–66). The three isoforms are generally present simultaneously, according to several analyses of human clinical isolates, but BFT-1 is slightly the most abundant. For further information, the amino acid sequences of BFT-1, BFT-2 and BFT-3 proteins are indicated in **Table S2, Annex III**.

	Position	BFT-1	BFT-2	BFT-3
Prodomain	32	A	T	A
	102	S	S	N
	169	D	D	Y
	170	I	L	P
	177	I	I	V
Catalytic domain	228	N	S	N
	232	I	V	V
	257	Y	F	Y
	260	F	L	L
	270	S	S	A
	275	D	N	N
	277	K	D	D
	281	E	D	D
	289	S	A	A
	312	K	N	N
	316	M	I	I
	319	F	L	L
	320	N	D	N
	331	E	K	K
	357	E	R	N
	359	T	A	A
	361	N	D	D
	362	S	P	P
	368	A	S	S
	369	T	K	K
	370	F	Y	Y
	375	S	F	F
	380	K	E	K
	383	D	Y	D
	384	I	R	I
	393	A	I	I

Table 1. Amino-acid variability between the three BFT isoforms. Table adapted from (67).

In some organisms, virulence genes are found in unique chromosomal loci, called pathogenicity islands (PAIs), which possess at least two virulence genes and have a different G+C content than the host chromosome, suggesting horizontal acquisition of sequences from a foreign organism (68). In the case of *B. fragilis*, the three BFT isoforms are encoded by a chromosomal PAI that is only present in ETBF, while is absent in NTBF strains (42). Several studies have revealed that there are five distinct patterns of hybridization, three of which predominated, and the strains with these genetic patterns were called pattern I, II and III *B. fragilis* strains (**Figure 7**) (69). In the case of ETBF, all strains belong to pattern I, due to their chromosomes possessing a 6-kb DNA region, not found in NTBF strains, which contains the *bft* gene and the *mpII* gene (metalloprotease II gene, which is a second putative virulence gene) (70). This 6-kb DNA region was termed *B. fragilis* pathogenicity island (BfPAI), which is flanked by genes encoding putative mobilization proteins, also called conjugative transposons (CTn) (42,69,70). Thus, this *bft* gene may be transmitted from ETBF to NTBF *B. fragilis* strains by lateral gene transfer (69,71,72). Some studies have found that the PAI and/or its flanking segments are present in 57% of the blood culture isolates analyzed. Concretely, the 19% of isolates had both elements, while the 38% of them had just the flanking segments (73).

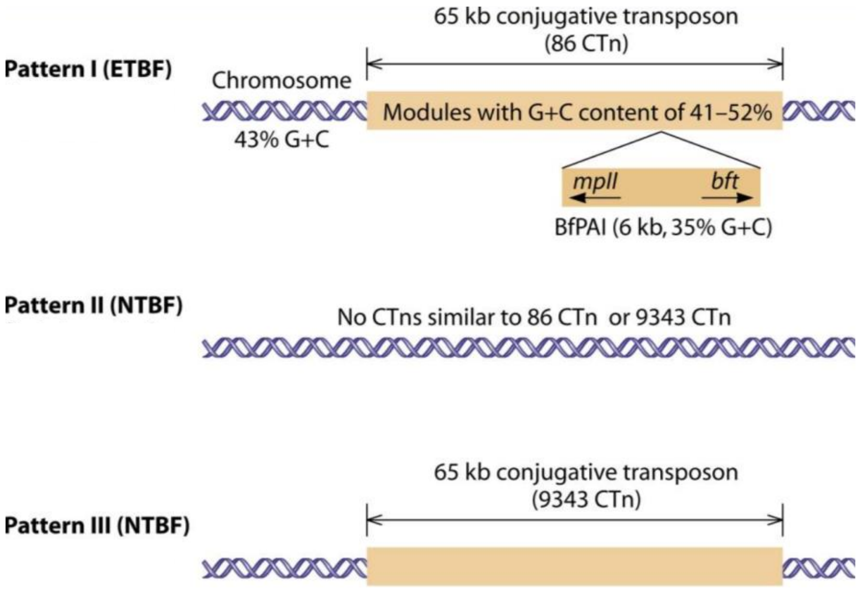

65 kb conjugative transposon
(86 CTn)
Pattern I (ETBF)
Chromosome
Modules with G+C content of 41–52%
43% G+C
mplI
bft
BfPAI (6 kb, 35% G+C)
Pattern II (NTBF)
No CTns similar to 86 CTn or 9343 CTn
65 kb conjugative transposon
(9343 CTn)
Pattern III (NTBF)

Figure 7. Different patterns of *B. fragilis* strains, corresponding to the different molecular types. Pattern I: ETBF strains with at least one 65-kb conjugative transposon (86 CTn). Pattern II: strains lack 86 CTn and 9343 CTn. Pattern III: NTBF strains with at least one 65-kb conjugative transposon (9343 CTn). Figure from (42).

1.2.1.3.2. *B. fragilis* toxin: structure

Regarding the BFT structure, the initial translation product is a preproprotein of 397 amino acids, which contains a 18-residue signal peptide at the N-terminus, followed by a proprotein which consists of a ~170-residue prodomain (PD), flanked by flexible segments, and a ~190-residue catalytic domain (CatD) (**Figure 8**) (67). The PD is cleaved at an Arg-Ala site, corresponding to amino acids 211 and 212, in order to become the mature BFT (63,64). The CatD has a zinc-binding consensus motif (HEXXHXXGXXH), and a Met residue close to this motif, being two features typical for zinc metalloproteases (MPs) termed metzincins. In this motif there are three histidines that coordinate the catalytic zinc ion (BFT has 1 g-atom of zinc per toxin molecule). Based on the HEXXH motif and other studies, BFT is classified as a zinc-dependent MP toxin (42,74). For instance, other related zinc-dependent MPs known to be important in human disease include tetanus, botulinum, and anthrax toxins (75).

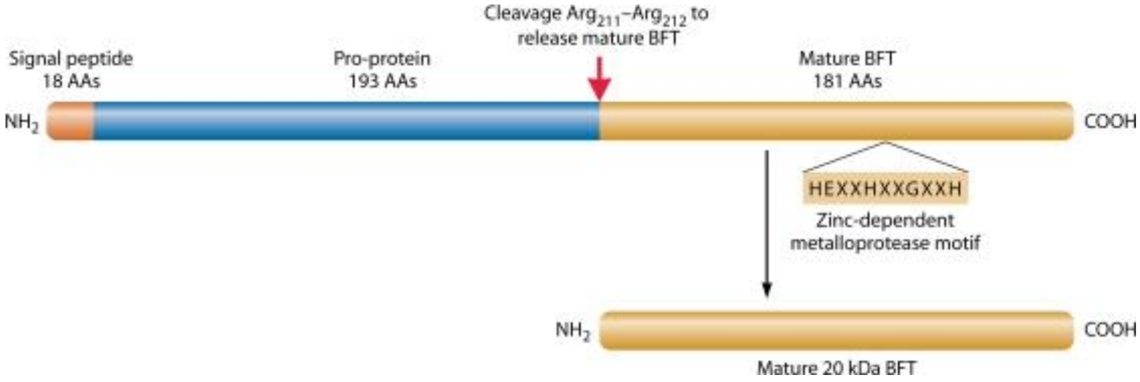

Cleavage Arg211–Arg212 to release mature BFT
Signal peptide 18 AAs
Pro-protein 193 AAs
Mature BFT 181 AAs
NH2
COOH
HEXXHXXGXXH
Zinc-dependent metalloprotease motif
NH2
COOH
Mature 20 kDa BFT

Figure 8. Scheme of the structure of *B. fragilis* toxin. Each of the three BFT isoforms consists of three protein domains: the signal peptide, the proprotein, and the mature toxin. In the image, at the beginning, the signal peptide is shown in orange. Then, the proprotein in blue, ending with the cleavage site (R^{211}-A^{212}) that releases mature BFT. Finally, the 181 amino acids in yellow represent the mature BFT protein, with the zinc-dependent motif (HEXXHXXGXXH), commonly present in zinc MPs. H, histidine; G, glycine. Figure from (42).

In a previous study (67) it was found that the PD plays an important role in both the latency and folding stability of the CatD. The protein is active only when the protein is in the mature form, i.e., CatD without the PD, indicating that the PD inhibits the proteolytic activity of BFT, which is crucial to is biologic activity (76). Besides, the PD has no significant sequence similarity to any known protein, but shows high structural similarity to adamalysins/ADAMs, which are only described in eukaryotes. Then, the limited identity existing between eukaryotic matrix MPs and BFT led to the hypothesis that BFT may be an ancestor of host matrix MPs (77). Then, it should have been co-opted by *B. fragilis* through a horizontal gene transfer mechanism from an eukaryotic cell (67).

This Doctoral Thesis has focused on the study of the isoform 3 (proBFT-3 as the immature form, and BFT-3 as the active one). The structure of proBFT-3 was obtained by Dr. Xavier Gomis-Rüth (67), a collaborator from the Molecular Biology Institute of Barcelona (IBMB), some years ago, and it was then deposited in the Protein Data Bank (PDB) server with the code 3P24 (**Figure 9**) (for further information see: https://www.rcsb.org/structure/3p24).

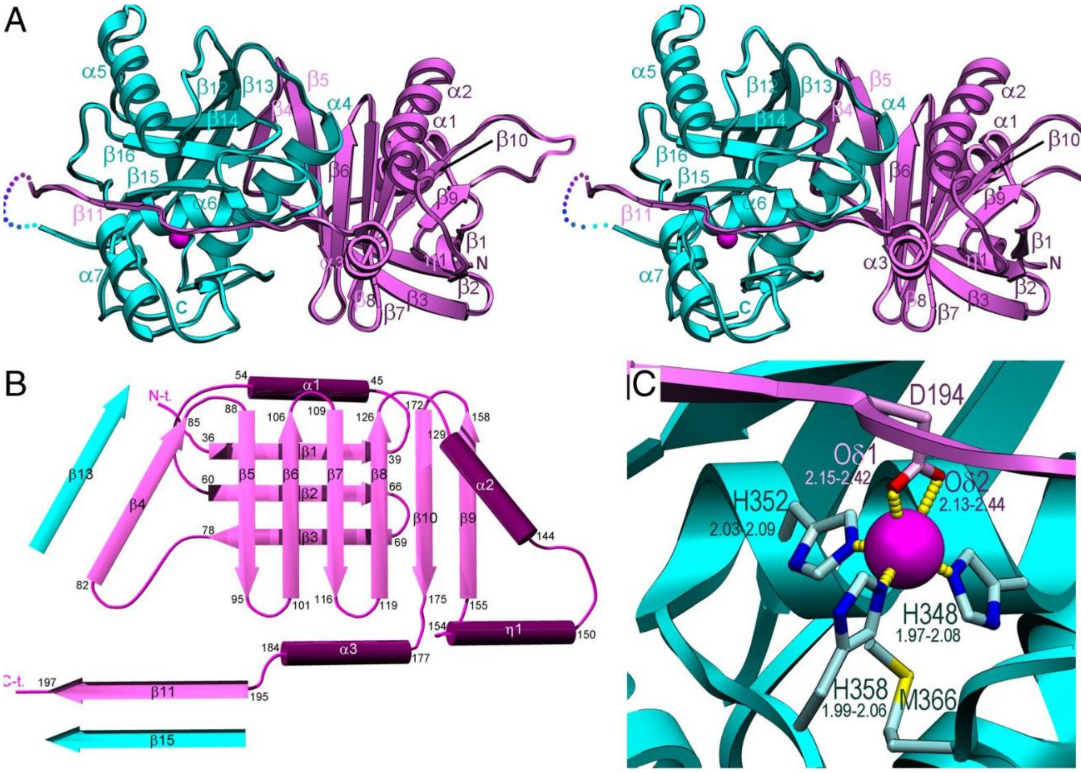

A
α5
β12
β13
β5
β4
α4
α2
α1
β10
β14
β16
β15
β6
β9
β11
α6
β1
α3
η1
N
α7
C
β8
β3
β2
β7
B
N-t.
α1
54
45
88
106
109
126
172
158
85
36
β1
39
129
β13
β4
60
β5
β6
β7
β8
66
α2
β2
β10
β9
78
β3
69
144
82
95
101
116
119
175
155
154
η1
150
184
α3
177
C-t.
197
β11
195
β15
C
D194
Oδ1
2.15-2.42
Oδ2
2.13-2.44
H352
2.03-2.09
H348
1.97-2.08
H358
1.99-2.06
M366

Figure 9. Overall structure of proBFT-3. (A) Richardson-type plot of proBFT-3, α-helices as ribbons and β-sheets as arrows. The PD is shown in magenta and the CatD in cyan, which is displayed approximately with the view into the active site cleft (the standard orientation characteristic for MPs). (B) Topology of the proBFT-3 PD that displays secondary structure with β-sheets (arrows) and α-helices (rods). (C) Zinc-binding site of proBFT-3, where the zinc is represented in magenta and it is being coordinated with three histidines of the CatD (H^{348}, H^{352}, H^{358}) and one Asp from the PD (D^{194}). Figure adapted from (67).

As previously mentioned, the BFT-3 has a zymogenic structure, which repressed its toxin activity through the self-inhibition of the PD, by preventing the access of substrates to the active site cleft. More precisely, the PD obstructs the access through the extended conformation of its C-terminal segment, since it is laid out across the entire CatD front in an opposite orientation to the substrate. In terms of latency, the segment which comprises α3, Lα3β11 and β11 plays a crucial role due to the fact that it traverses the front of the CatD from the right to the left. This segment forms a parallel β-sheet interaction with the cleft through the strand β11 with the "upper-rim strand" β15 of the CatD on the nonprime side (**Figure 10**) (67).

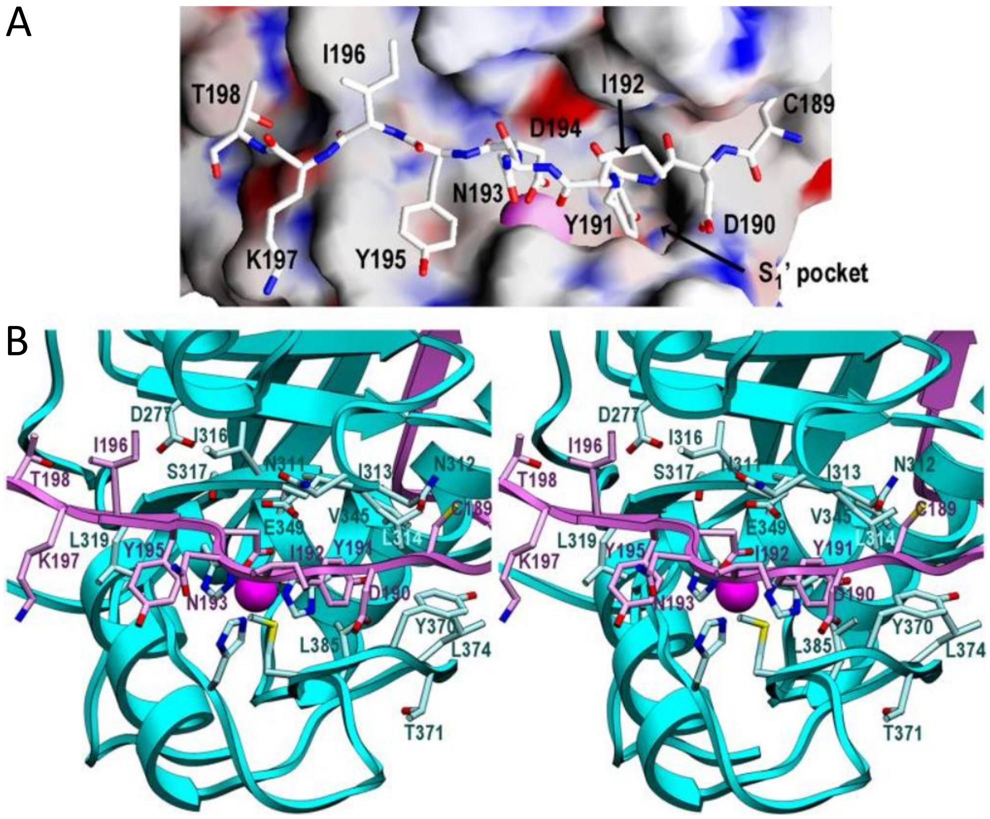

A
I196
T198
I192
C189
D194
N193
Y191
D190
K197
Y195
S1' pocket
B
D277
I316
I196
S317
N311
I313
N312
T198
E349
V345
C189
L319
Y195
I192
Y191
L314
K197
N193
D190
Y370
L385
L374
T371

Figure 10. Active site cleft of proBFT-3. (A) Active site cleft of BFT-3 superimposed with its Connolly surface colored basing on the electrostatic potential. The PD segment C189-T198 is shown as a stick model, which is in the pocket of the CatD preventing the access of substrates. (B) A closer view of the Figure 9A showing the active site environment, including the pocket and the involved residues. All residues except those already tagged in Figure 9C are labeled. Figure adapted from (67).

In terms of its stability, the BFT protein is considered stable at room temperature and below, but a rapid autodigestion occurs above 37 °C. Concretely, there is a stepwise autolytic processing of proBFT-3 at 37 °C over time in which various intermediates can be observed, and finally there is a stable 20 kDa form (**Figure 11**). Studies indicated that autolysis only may occur *in vitro* in the purified proBFT, while heterolytic activation is the mechanism that occurs *in vivo*. There are some candidates responsible for this heterolytic activation *in vivo*, and the most likely candidate is trypsin, which is expressed both in stomach and small intestine, as well as in cancer cell lines derived from stomach and colon (67).

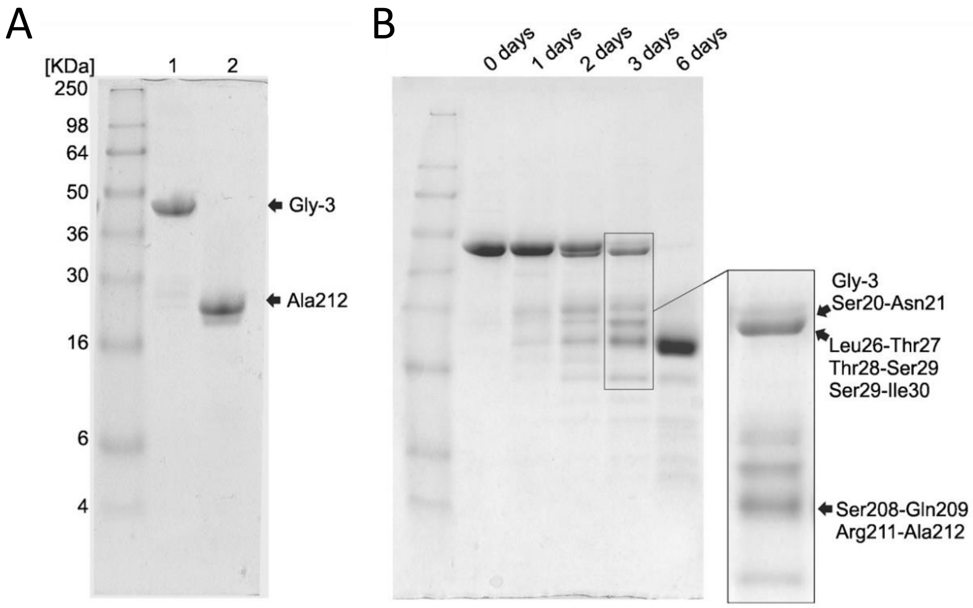

A
[KDa]
1
2
250
98
64
50
36
30
16
6
4
Gly-3
Ala212
B
0 days
1 days
2 days
3 days
6 days
Gly-3
Ser20-Asn21
Leu26-Thr27
Thr28-Ser29
Ser29-Ile30
Ser208-Gln209
Arg211-Ala212

Figure 11. ProBFT-3 activation studies *in vitro*. (A) SDS-PAGE of purified proBFT-3 (lane 1) and trypsin-activated BFT-3 (lane 2) with the respective N-terminal residues. (B) Autolytic cleavage of proBFT-3 at 37 °C over time. The cleavage sites after 3 days are shown in the framed close-up view. Figure from (67).

Due to the aforementioned, under certain *in vitro* or *in vivo* situations the protein is present as inactive protein (proBFT-3), active protein (BFT-3), or a mixture of both forms depending on those conditions. Therefore, throughout this Doctoral Thesis we distinguish the immature form (proBFT-3) from the mature form (BFT-3) at each moment, but sometimes the distinction will be somewhat blurred; in those cases, the protein will be indicated with the name that is considered most appropriate.

1.2.1.3.3. *B. fragilis* toxin: functions and implications

B. fragilis toxin can destroy the zonula adherens tight junctions in intestinal epithelium, by causing cell damage modifying the target cell surface via an unknown colonic epithelial cell (CEC) protein receptor. As a result, morphological appearance is observed in cell cultures as shown in **Figure 12** (44,75).

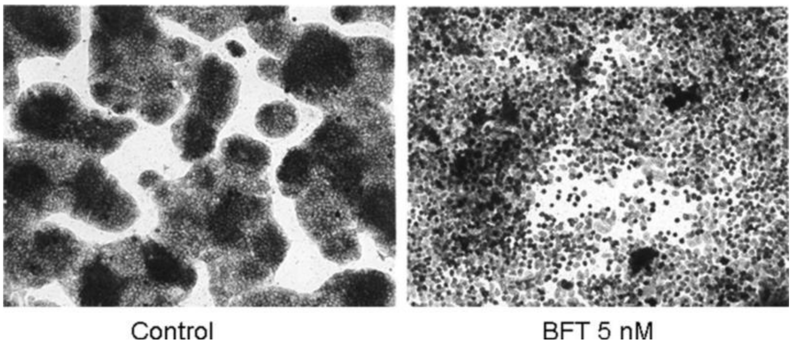


Figure 12. Morphological effect of BFT on HT-29/C1 cells *in vitro*. After treating HT-29/C1 cells with 5 nM BFT, these cells undergo morphological changes, including cell rounding and dissolution of cell clusters. Figure from (42).

Briefly, BFT provokes an increase in both permeability and chloride secretion in intestinal epithelial cell (IEC) monolayers, rearrangements related to the actin of the cytoskeleton and loss of tight junctions; actions known to promote diarrhea, mucosal inflammation and therefore colon inflammation and enhancement of metastatic potential/carcinogenesis (31,53,78,79). After the interaction of BFT with the CEC receptor, the cellular effect is quite fast and it rapidly triggers subsequent events. Among others, BFT stimulates multiple host cell changes including: reduction of the barrier function by inducing the cleavage of the zonula adherens protein E-cadherin, triggering β-catenin-dependent nuclear signaling, activation of the nuclear factor kappa B (NF-κB), induction of the proto-oncogene protein *c-Myc* expression, induction of expression and secretion of cytokines such as the interleukin-8 (IL-8), induction of DNA damage and cellular proliferation (**Figure 13**) (76,80).

As mentioned above, one of the processed proteins by BFT is the E-cadherin, which is a 120 kDa glycosylated Type I transmembrane protein and the principal protein of the zonula adherens tight junctions in the intestinal epithelium. E-cadherin has several important roles related to cellular signaling, proliferation and differentiation, but, above all, to the formation and integrity of the intercellular adhesions junctions (47,81,82).

The E-cadherin domain which is associated with β-catenin is the cytoplasmic one, and β-catenin is in turn tethered to both α-catenin and actin. This protein complex is important for mediating calcium dependent homotypic cell-cell junctions in several tissue types, thus maintaining epithelial polarity and integrity (83).

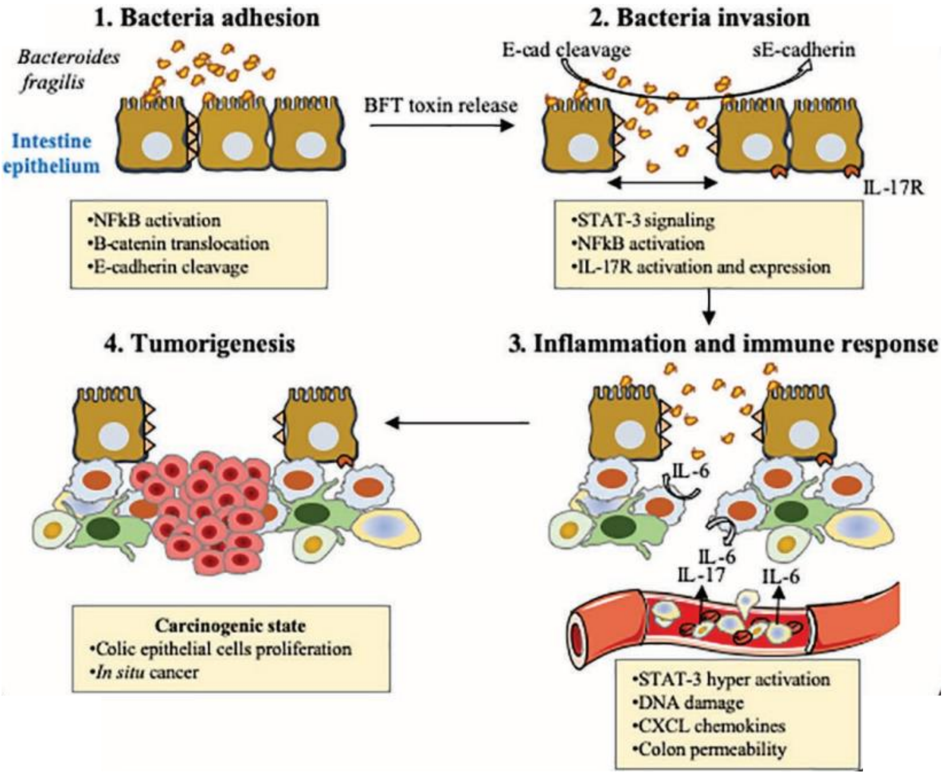

1. Bacteria adhesion
Bacteroides fragilis
Intestine epithelium
BFT toxin release
2. Bacteria invasion
E-cad cleavage
sE-cadherin
IL-17R
•NFkB activation
•B-catenin translocation
•E-cadherin cleavage
•STAT-3 signaling
•NFkB activation
•IL-17R activation and expression
4. Tumorigenesis
3. Inflammation and immune response
IL-6
IL-6
IL-17
IL-6
Carcinogenic state
•Colic epithelial cells proliferation
•In situ cancer
•STAT-3 hyper activation
•DNA damage
•CXCL chemokines
•Colon permeability

Figure 13. Cellular host E-cadherin as target during *B. fragilis* infection. In the image, the hypothetical model of tumorigenesis associated with ETBF strains. Figure adapted from (84).

It was previously thought that BFT cuts directly the E-cadherin ectodomain, but there is no data to support such direct interaction. In fact, although BFT is a proteolytic toxin, the data do not support cleavage of E-cadherin directly by BFT, nor does BFT should modify protease-activated receptors (75). The first step of the currently proposed mechanism for the E-cadherin cleavage is the binding of BFT to a host CEC receptor. Then, BFT rapidly induces the scission of the 80 kDa extracellular E-cadherin ectodomain, followed by a stimulating host cell γ-secretase activation that, either alone or complexed with BFT, processes the intracellular cell-associated E-cadherin, yielding sequentially 33 kDa and 28 kDa fragments (**Figure 14**) (75,76,81).

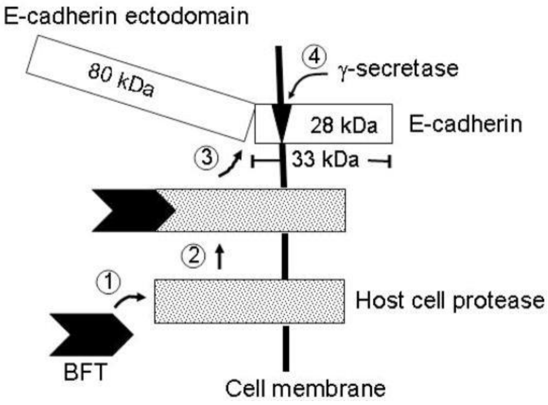

E-cadherin ectodomain
80 kDa
④
γ-secretase
28 kDa
E-cadherin
③
33 kDa
②
①
Host cell protease
BFT
Cell membrane

Figure 14. Proposed model of BFT mechanism of action. Figure from (81).

BFT-mediated cleavage of E-cadherin initiates a multi-step inflammatory cascade (**Figure 15**). More concretely, β-catenin is released to the cytoplasm, promoting then β-catenin nuclear localization and stimulation of β-catenin–T Cell factors-dependent cellular proliferation (80,81). This scission also triggers Wnt signaling, involved in the regulation of cell proliferation, and thus reducing cell adhesion and favoring cancer cell migration due to weakened tight junctions (43). When BFT is attached to the receptor, it also triggers CEC signal transductions which involve tyrosine kinases (TK) and mitogen-activated protein kinases (MAPK) (42). The NF-κB signaling, related to anti-apoptotic effects, is also activated not only in the epithelial cells which are transformed, but also in myeloid cells that contribute to inflammation. Then, there is an IL-17 dependent NF-κB activation, which contributes to a mucosal gradient of CXC chemokines implicated in the initiation pro-tumoral myeloid cell infiltration from the distal part of the colon (36,51,84). The NF-κB activation also induces the nitric oxide synthase which leads to ROS production, such as nitric oxide, that induces DNA cellular damage producing thus oncogenic alterations (43,50). IL-6 is also implicated in the inflammation produced by NF-κB, which induces the Stat3 pathway, a procarcinogenic signaling pathway leading to the activation of proliferative, antiapoptotic and proangiogenic genes involved in cancer growth (51,85). The mentioned pathway is responsible for blocking the generation of antitumor immune responses by organizing the immune microenvironment of tumors (86). Besides, the nucleus translocated β-catenin produced by E-cadherin cleavage binds to transcriptional activators, such as the T-cell factor (TCF), to enhance the transcription of *c-Myc* oncogene, thus increasing cellular proliferation. *c-Myc* mutations or overexpression lead to carcinogenesis (50,80). The epithelial cell injury produced by BFT exposure stimulates the production of the proinflammatory cytokine IL-8, a potent chemoattractant of polymorphonuclear cells (87).

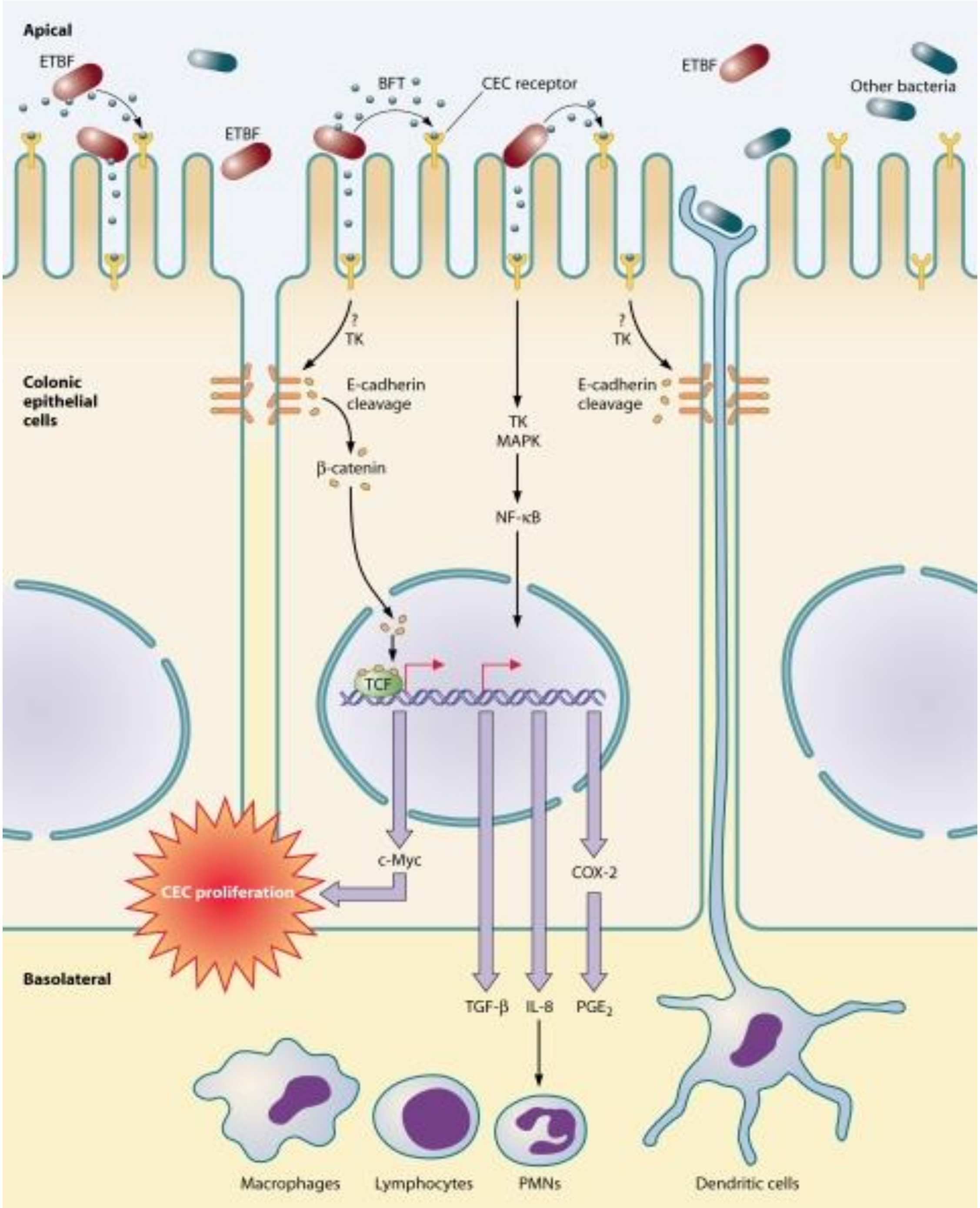

Apical
ETBF
ETBF
BFT
CEC receptor
ETBF
Other bacteria
Colonic epithelial cells
?
TK
?
TK
E-cadherin cleavage
E-cadherin cleavage
TK
MAPK
β-catenin
NF-κB
TCF
c-Myc
CEC proliferation
COX-2
Basolateral
TGF-β
IL-8
PGE₂
Macrophages
Lymphocytes
PMNs
Dendritic cells

Figure 15. Model of the mechanism of action of BFT in ETBF strains. Figure from (42).

1.2.1.4. *B. fragilis* treatment in the context of antimicrobial resistance

For decades, antibiotics have been the "wonder antimicrobial drugs" to combat infections because only spontaneous cures, surgical drainage antiseptics, arsenicals or silver compounds were available to treat serious bacterial infections before they were employed (88). In fact, since the discovery of penicillin in 1928 (89), they have revolutionized the medicine for more than 60 years, and it is estimated that the widespread commercialization and development of antibiotics has increased human life expectancy in developed countries by 30 years (90,91). Besides, apart from its widely known use for therapeutic purposes, they have been used in many other industries such as agriculture and animal husbandry (88,92).

Many of these antibiotics have activity against a broad spectrum of bacteria, from infections caused by Gram-positive to Gram-negative bacteria. As a consequence of this low specificity, there are few antibiotics commonly known as "ideal antibiotics", which means that in general an antimicrobial agent is not able to kill or inhibit the specific growth of a harmful bacterium, without affecting the rest of the beneficial microbes with which it shares the site of infection (gut/skin flora) (91). In this context, broad-spectrum antibiotics could remove enteropathogens from the gastrointestinal tract, but they also may affect the commensal flora, which is beneficial for the host. Thus, the problem is that, in the absence of this flora, opportunistic microorganisms may colonize the intestine and lead to severe gastrointestinal diseases. This fact, together with the continuous increase in the demand for antibiotics across many sectors, and the irresponsible overuse of them, has greatly contributed to the emergence of resistant strains (93). In addition, there is no record of discovery of any new class of antimicrobial agents since 1987 (94). Thus, this is the scenario for antimicrobial resistance (AMR) arising as a global health emergency for the last years. In fact, WHO has long been warning of the enormous problem of multidrug-resistant bacteria, which are causing (and will cause) more deaths worldwide. Specifically, it is predicted that we are going to be in a post-antibiotic era that will result in more frequent infections and even small injuries could result in death if we fail to act against antibiotic resistance (92). On the basis of specific data, it is estimated that drug-resistant infections kill more people than HIV/AIDS (864,000 deaths) or malaria (643,000 deaths). Specifically, in 2019, 4.95 million deaths were associated with bacterial AMR, including 1.27 million direct deaths attributable to bacterial AMR (95). Regarding these data, there is no doubt that AMR poses a potential threat to human health worldwide. Moreover, based on several studies, the situation is expected to worsen in the long run if we do not act correctly. It is estimated that up to 10 million people could die annually from AMR by 2050 (96,97). Therefore, in order to reduce AMR, making informed and location-specific policy decisions regarding control programs to prevent infections, the access to essential antibiotics, and the research for the discovery and development of new antibiotics, as well as vaccines, are crucial.

In the context of AMR, *Bacteroides* spp. presents “the most antibiotic-resistant mechanism and highest resistant rates of all anaerobic pathogens" (31). Then, a suitable use of effective treatments is essential for a favorable clinical outcome in infections caused by *B. fragilis*. This bacterium is commonly exposed to different antimicrobial and antibiotics, and as in other cases of bacterial resistance to antimicrobials, multidrug-resistant strains have emerged to several classes of structurally unrelated antibiotics. *B. fragilis* evolved from being susceptible to become resistant to a broad spectrum of antimicrobial agents. Concretely, some ETBF strains can be resistant to several antibiotics such as penicillin, ampicillin, clindamycin, tetracycline and metronidazole (42,67). Analyzing the antibiotic resistance pattern of 78 *B. fragilis* isolates, it was observed that the highest rate of resistance was for penicillin G (100%) followed by tetracycline (74.4%), clindamycin (41%) and cefoxitin (38.5%). Besides, all isolates were susceptible to metronidazole (38). Some of the known mechanism of resistance found for this bacteria are: enzymatic inactivation (for β-lactams), efflux pumps-mediated resistance (for β-lactams, metronidazole, quinolones, tetracyclines), and mutations in gyrase (*gyr*A) and topoisomerase IV (*par*C) (for quinolones) (36). The most antibiotic-resistant genes identified in *B. fragilis* were tetracycline resistance protein (*tetQ)*, cephalosporinase (*cepA*) and erythromycin ribosomal methylase (*ermF*) (38). Apart from the fact that treating infections caused by *B. fragilis* is currently a challenge, these bacteria have the potential to act as a reservoir of antibiotic-resistant genes, resulting in their horizontal transfer to other components of normal bacterial flora through integrated transposons, integrated genetic elements, as well as conjugative plasmids (98).

The current treatment for *B. fragilis* infections with antibiotics depends on each country's health systems; but, to summarize, β-lactams (such as cefoxitin, which is a cephalosporin) co-administered with β-lactamase inhibitors, carbapenems (such as imipenem and meropenem), clindamycin, and metronidazole are frequently prescribed (the latter two often in combination with the new generation of fluoroquinolones) (38,98,107–110,99–106). Metronidazole is an *in vitro* highly active agent against *B. fragilis* isolates, and that is why it remains in the first-line of antimicrobial empirical therapy (38).

Importantly, there is an important consideration to be done: administering a broad-spectrum antibiotic may tackle the BFT toxin, but it may affect as well the beneficial flora. In this way, orally administered broad-spectrum antibiotics may eliminate enteropathogens from the gastrointestinal tract but they also affect the commensal flora that humans need. Thus, opportunistic microorganisms may colonize the intestine in the absence of this beneficial flora, and lead to severe gastrointestinal diseases. That is why, there is an urgent necessity to develop specific antimicrobial treatments or to redesign the existing ones, in order to achieve a specific inhibition of BFT, to combat ETBF-mediated pathogenicity, without disturbing the commensal microbiota (38,98,108–110,100–107). Accordingly, there is an indisputable need for better understanding ETBF targets, such as BFT-3, in order to identify highly specific antimicrobials to tackle them (42,67,111).

1.3. Animal models for studying *B. fragilis* infections

The first works using animal models in the study of human anatomy and physiology date back to the 6th century, when it was determined that the brain is the seat of intelligence and sensory integration using dogs (112). Since this date, their use has been exponentially growing in the pursuit of medical and scientific knowledge, leading to significant contributions to the current medical understanding (113). This fact was possible due to the “comparative medicine” concept, which was based on the idea that other animal species share physiological, behavioral or other characteristics with humans (112). That is why the study of human diseases simulated in animal models has played important roles in the understanding many pathologies. In this context, murine models have been the “gold standard” for studying microbial infections over the past decades (114,115). Nowadays, around 20 million animal subjects are employed in biomedical research worldwide, and most of them are mouse and rat models (113).

In the field of microbiology, the use of these animal models has enabled to understand the pathogenesis and treatment of infections caused by different types of bacteria. Animal models are needed not only for studying pathogenic processes, but also for determining whether a therapeutic agent or immunologic intervention is likely to properly work when applied to human population (116,117). The use of animal models to simulate and operate under *in vivo* conditions is required to stablish the role of pathogens during serious infections in order to better understand which is the most appropriate antibiotic therapy. Besides, a key point of the assays in animal models is to tailor the therapeutic doses as well as intervals of administration to match the concentration peak and trough levels for the same antibiotics in humans. In these kind of studies, it is also possible to evaluate the potency, efficacy and toxicity of drugs selected in *in vitro* studies to combat infections (116).

In the case of *B. fragilis* studies in animal models, several assays for different purposes have been carried out. Some examples are detailed below.

Using rats, it was stablished that, despite the fact that *B. fragilis* is relatively a minor component of the intestinal microflora, is the most common Gram-negative anaerobe isolated from humans with intra-abdominal infections. These studies indicated that unique virulence factors could play important roles in the prominence of microorganisms (116).

Several studies have been performed to determine possible synergies between bacteria involved in intra-abdominal sepsis. Many of these observations have been carried out in animal models of rats, guinea pigs and mice, using wound infection models. For example, a result of this included the synergy between *E. coli* and *B. fragilis* in provoking purulent infections (118,119).

Apart from the synergies that *B. fragilis* establishes with other organisms, the presence of a capsular polysaccharide, not detected for other phenotypically similar organisms, allows it to cause intra-abdominal abscesses in rats without the presence of a synergistic partner (116).

In a study using a mice animal model for human multiple sclerosis, it was showed that oral administration with the purified zwitterionic capsular polysaccharide A of nontoxigenic *B. fragilis* strains can protect, prophylactically and therapeutically, against central nervous system demyelinating disease (120).

However, despite the enormous advances that have been achieved with the use of animal models in research, the use of them has been called into question by the entire scientific community during the last years due to concerns about their clinical validity and applications, and they have been also questioned by the general population, above all due to ethical concerns. For these reasons, there is widespread thinking about reducing the use of animals for research (114,121,122). Furthermore, the use of such animals requires the maintenance of sufficient numbers of them to achieve statistically relevant data, leading to expensive as well as laborious protocols.

Then, in response to public concerns about the use of vertebrate animals in research, the interest in invertebrate models has increased over the past several decades. As result, for many diseases and conditions, invertebrates are being evaluated and recognized as models in research. Using them, an enormous quantity of discoveries in almost every area of biology and medicine have been achieved, for example from embryonic development to aging processes (115,123). Then, due to the fact that a bridge between *in vitro* studies and studies in mammals is clearly necessary, alternative host models such as *Acanthamoeba castellanii* (amoebae), *Artemia salina* (brine shrimp), *Caenorhabditis elegans* (roundworm), *Danio rerio* (zebra fish), *Drosophila melanogaster* (fruit fly) and *Galleria mellonella* (greater wax moth), are being employed (124,125).

1.3.1. *G. mellonella* waxworm: an overview of its increasing importance

Since 2016 the use of the *G. mellonella* waxworm as an alternative infection model organism, for studying many pathogens as well as searching for new drugs, has increased by 42% (125,126).

This increase has been possible because several studies have shown a positive correlation between results obtained in *G. mellonella* and mammalian models (127–129).

1.3.1.1. General characteristics of *G. mellonella*

The *G. mellonella* insect, also called honeycomb moth or greater wax moth, is a member of the Galleriinae subfamily within the Pyralidae family of the Lepidoptera order (**Figure 16**) (125,130). *G. mellonella* has a wide geographic distribution and the natural habitat of this invertebrate is beehives, where they have the possibility of feeding on pollen, beeswax and honey. In their natural environment, there are several microbes for which these insects have evolved immune response; therefore, their amenability to infection, as well as their ability to produce a defense response, make *G. mellonella* an interesting host in the study of microbial pathogenesis (131).

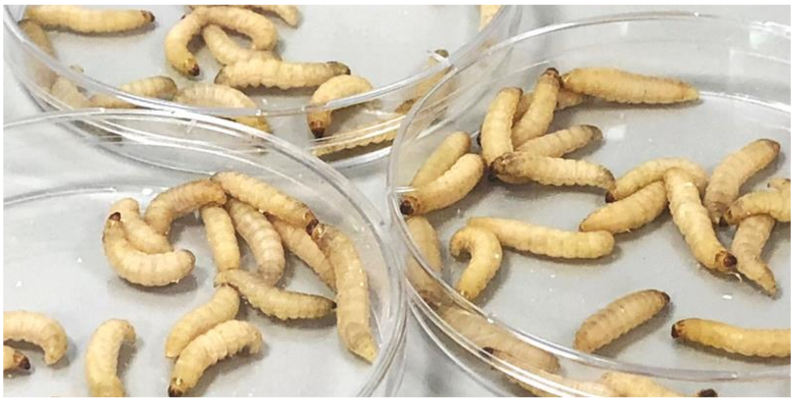


Figure 16. *G. mellonella* larvae appearance.

1.3.1.2. Life cycle of *G. mellonella*

Regarding to the developmental stages, *G. mellonella* insects develop through four distinct life stages: egg, larvae, pre-pupae/pupae and adult insect (**Figure 17**) (125,132). Complete metamorphosis depends on both biotic and abiotic factors, and among them, the most important are competition for food, diet quality, temperature, cannibalism and relative humidity. Depending on environmental conditions, such as temperature and food supply, the entire life cycle takes approximately 40 days (130,133–135).

The first phase corresponds to eggs, which have a spheroidal shape, are white to light pink, and take between 5-8 days to develop and became larvae (at a temperature comprising between 24-27 °C), while the development into larvae lasts between 3-30 days (if the temperature is 10-16 °C). Therefore, the development into larvae is temperature dependent (126). Next, during 6-7 weeks, the insects are larvae with a creamy-white color, with a reddish head and they measure from 1-23 mm (133,134). During this period, there are some molting stages, and they spin silk threads across all stages, but only in the last instar they spin a cocoon (133,134). A larva consists of a number of segments, which are divided into 3 anatomical parts: head, thorax, and abdomen. They have six legs founded at the thoracic part and some prolegs presented on the abdominal area (126). The moment in which the larvae stop feeding but they are still motile, corresponds to the intermediate stage of the pre-pupae (136). In the following stage, the larvae became pupae, where they turn from white to yellow, then to brown and finally take a dark reddish brown and are immobilized in cocoons. From pupae to moth stage about 1-8 weeks are necessary, and during this time, the insect does not eat. Finally, the insects develop into adult moths, which are reddish brown and pale cream color, active during the night because they are sensitive to light, and are able to lay 50-150 eggs. As in the pupae stage, adult moths do not feed. Male moths are slightly smaller and lighter in color than females, and they live 21 days whereas females live 12 days (126,130,132–135). The short life cycle of these worms facilitates the large-scale studies (137).

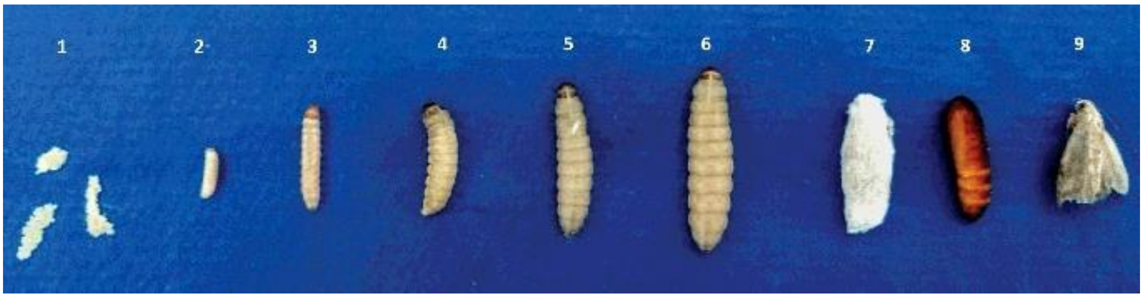


Figure 17. Different developmental stages of *G. mellonella*. 1. Eggs; 2. 10-day-old larvae; 3. 20-day-old larvae; 4 and 5. 25-35-day-old larvae; 6. 40-day-old larvae (last larval stage); 7 and 8. Pre-pupae and pupae; 9. Adult moth. Figure from (132).

As previously mentioned, during the development of the insect, the temperature is a crucial factor, and the optimum temperature averages between 29-33 °C. Moreover, larvae can survive at 37 °C, which is the mammalian physiological temperature (125,130,138).

1.3.1.3. Reasons for the increased use of *G. mellonella*

Regarding *G. mellonella* as animal model, there are several important aspects that have contributed to its increased use in the last years.

A key point is the fact that the larval immune system of *G. mellonella* has remarkable functional and structural similarities to the innate immune response of mammals, including phagocytosis and the production of antimicrobial peptides and both ROS and nitrogen species (**Figure 18**)

(139,140). Among other important aspects, the insect cuticle is a barrier to infection much as is the mammalian skin, the microvilli in the midgut of *G. mellonella* contains microbes that resemble those found in the intestinal microvilli of mammals, and several similarities have been described between epithelial cells from mammalian digestive tract and those from larvae (126,141,142). Besides, these waxworms possess hemolymph, which is analogous to mammalian blood and contains immune cells called hemocytes. The concentration of hemocytes fluctuates during worm life, and it is affected, for example, by the stress caused by microorganisms (143). These cells can be compared to neutrophils in mammals in terms of physiological functions, due to the fact that they have the ability to phagocyte and kill pathogens by producing superoxides, and they are also implicated in encapsulation, nodulation, clotting and melanization processes (126,140). At least, six different type of hemocytes have been identified within *G. mellonella*, which are called prohemocytes, plasmatocytes, granular cells, coagulocytes, sphelurocytes and oenocytoids (144). In addition, *G. mellonella* produce soluble effector molecules, which are involved in the humoral response. Some of them are complement-like proteins (opsonins), melanin, antimicrobial peptides (AMPs) and insect metalloproteinase inhibitors (126,145). At least, four opsonins classes have been described, which are apolipophorin-III (apoLp-III), peptidoglycan recognition proteins (PRGPs), cationic protein 8 (GmCP8) and hemolin, and they target bacteria cell-wall components such as lipopolysaccharide (LPS) or lipoteichoic acid (LTA) (122,146). The repertoire of AMPs consists of at least 20 peptides with a broad spectral microbial activity, so they are components that play the main role in *G. mellonella* innate immunity (147,148). Using different mechanisms, AMPs provoke the induction of bacterial cell lysis by ion leakage. Two types of AMPs have been described, called anionic and cationic antimicrobials (126,146). *G. mellonella* also produces a microbial-induced metalloproteinase inhibitor, called insect metalloproteinase inhibitor (IMPI), which is responsible for inhibiting thermolysin-like metalloproteinases from pathogens, being implicated in this way in the innate immune response (139,148,149). These insects also produce lysozymes, which are found in the hemolymph and act as antimicrobial proteins by hydrolyzing the peptidoglycan of bacteria, then modulating the larvae microbiota (139,148). Another important process implicated in the humoral response is the melanization pathway, which participates against bacteria or fungi in the antimicrobial activity (132,150). Basically, when an infection process occurs, there are some soluble effector molecules which bind to bacterial components such as LPS or LTA, provoking the induction of the release of prophenoloxidase by oenocytoids. Then, through a serine protease cascade, phenoloxidase (PO) is activated, which oxidizes the phenolic compounds into quinones and then they are metabolized into melanin, explaining the black spots in the infected larvae (**Figure 19**) (151,152). This melanization pathway reinforce the use of *G. mellonella* as animal model due to the fact that it is very easy to visually analyze the number of live and dead worms in an assay. Besides, due to the fact that *G. mellonella* do not have T and B lymphocytes neither produce antibodies, they do not possess adaptive immunity. However, despite it has not been clearly elucidated, a specific response called 'trained immunity', has been described. This kind of immune-like memory consists of an infection containment mediated by hemocytes and AMPs, and it is evidenced by a first infection at sub-lethal concentrations of a pathogen, followed later by the exposure to higher concentrations of the pathogen, resulting in more resistant larvae (126,153,154).

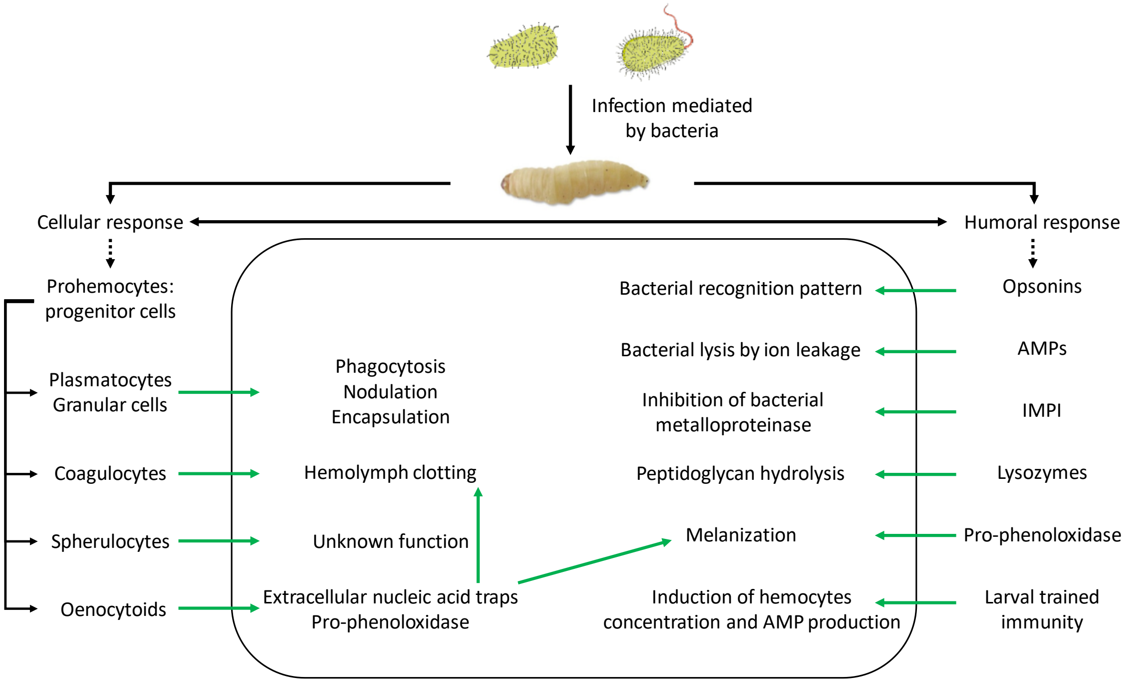

Infection mediated by bacteria
Cellular response
Humoral response
Prohemocytes: progenitor cells
Plasmatocytes Granular cells
Coagulocytes
Spherulocytes
Oenocytoids
Phagocytosis Nodulation Encapsulation
Hemolymph clotting
Unknown function
Extracellular nucleic acid traps Pro-phenoloxidase
Bacterial recognition pattern
Bacterial lysis by ion leakage
Inhibition of bacterial metalloproteinase
Peptidoglycan hydrolysis
Melanization
Induction of hemocytes concentration and AMP production
Opsonins
AMPs
IMPI
Lysozymes
Pro-phenoloxidase
Larval trained immunity

Figure 18. *G. mellonella* immune system activation after a bacterial infection. On the left, the cellular immune response; on the right, the humoral response; on the middle, the pathophysiological consequences (green arrows indicate activation).

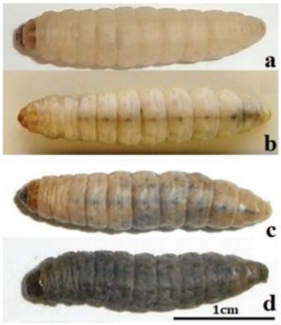


Figure 19. Color patterns of *G. mellonella* larvae after a bacterial infection. Untreated/control larvae show no melanization. Depending on the grade of melanization, the process can be observed as (a) no melanization, (b) slight, (c) moderate or (d) severe melanization. Figure adapted from (155).

In this context, the innate immune response of mammals is a vital component in the immune response to pathogenic infections and results obtained using insects show strong correlation with those obtained using mice. Thus, regarding this aspect, *G. mellonella* represents an effective bridge between *in vitro* studies and studies in mammals (122,124,140,156).

Another relevant aspect in the increase of the use of *G. mellonella* as animal model is that, unlike other non-mammalian model organisms, such as *C. elegans*, *D. rerio*, and *D. melanogaster* (157–159), *G. mellonella* worms can be incubated at 37 °C, allowing the study of clinically relevant human pathogens at a temperature that mimics that of humans, as well as studying temperature-dependent virulence factors, since temperature regulates the expression of a broad range of virulence factors (125,145). Besides, they are cheap and easy to obtain from commercial insect suppliers, and can be housed in large numbers to allow large-scale studies at low cost and without requiring expensive equipment for their maintenance nor requiring major adaptations to the infrastructure of laboratories (137,145). Furthermore, as insects, they are not currently subject to exhaustive ethical restrictions and welfare legislation as it occurs with other animal models, meaning that the barrier to entry for researches, who wish to move their studies into an animal model host, is lower than in other cases (145,160,161).

Infection of these worms is generally carried out on fifth instar insects, typically around 2 cm long with a mass of 180-250 mg (141). The most common method of infection is by an injection into the haemocoel through the last proleg of the insect (162), which is the main cavity in many invertebrates and where the circulatory fluid is located, previously mentioned as hemolymph (125). In addition, the small size of these worms facilitates the injection process.

Nevertheless, there are aspects that may affect the assays, because their used is limited after the delivery, their good health depends on several extrinsic factors such as the temperature (more affected in the hotter months), and transport conditions, and besides, there is a lack of standardization of the conditions, since there is a huge variability in the maintenance and treatment methods from one laboratory to another (162–164). For instance, infection methods vary between laboratories: the most common method of infection is by injection (with 10 µL inoculum into the haemocoel through the last proleg of the insect) (**Figure 20**) (131,162), but another options are force-feeding (141,165) or the topical application (131). But above all, the most limiting factor is learning to work with them, since learning how to perform the injections is a complex process, and the researcher must be careful not to suffer punctures.

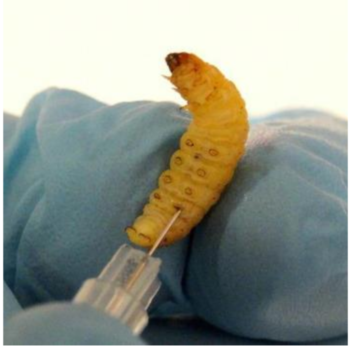


Figure 20. Procedure for inoculating *G. mellonella* larvae through an injection in the last left pro-leg in order to induce an intra hemocoelic infection.

A summary of the advantages and disadvantages of the use of *G. mellonella* as animal model is shown in the **Figure 21**.

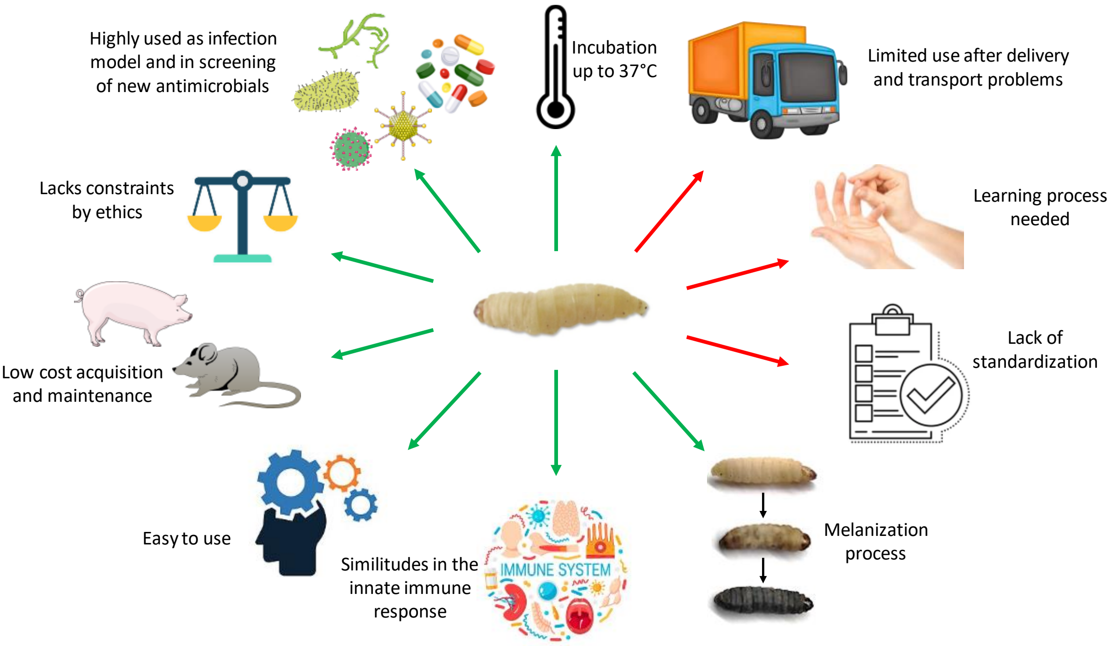

Highly used as infection model and in screening of new antimicrobials
Incubation up to 37°C
Limited use after delivery and transport problems
Lacks constraints by ethics
Learning process needed
Low cost acquisition and maintenance
Lack of standardization
Easy to use
Similitudes in the innate immune response
IMMUNE SYSTEM
Melanization process

Figure 21. Advantages (green arrows) and disadvantages (red arrows) of *G. mellonella* as animal infection model.

Due to all of the above, and in order to apply the '3Rs' principles (replacement, reduction and refinement) in animal experimentation, aiming to reduce the number of mammals, substituting in this way the vertebrate infection models used in research (145), as well as thanks to the fact that the genome of *G. mellonella* has recently been sequenced, its use as a model organism in the study of diseases caused by bacteria, fungi and viruses (166), and its use for the screening of new antimicrobials, has strongly grown (145).

1.3.1.4. Differences in *G. mellonella* management protocols: different methods for administration

As briefly mentioned above, infection methods vary between laboratories, but larvae are commonly injected with 10 µL inoculum. The most common method of infection is by injection into the haemocoel through the last proleg of the insect, but there is a considerable variability among methods for injection. One usual method is to immobilize the needle itself and then perform the injection once the waxworm is onto the needle. Another method is to immobilize the waxworms between the operator's fingers and place the needle into the insect's proleg, lifting the needle away from the operator with the insect attached, in order to confirm a suitable injection, before pushing the plunger on the syringe. Both of these injection techniques present a hazard to the researcher and can result in needle stick injury and possible infection (131,162).

Another option is to use a force-feeding model since many pathogens enter the host via the oral colonization. In this case, a dye should be added to the solution intended for force-feeding, such as a strong bromophenol dye, in order to remove any injured larvae and just select for those that contain the dye only within their gut (141,165).

There are other options such as the topical application, which is very simple since to the inoculum is placed on the larvae cuticle. This method is used more for fungi infections, where the fungi enter into the larvae by penetrating the insect cuticle. While facilitating the process, this delivery method has the disadvantage of an uncertain number of infecting cells (131).

1.3.1.5. *G. mellonella* as infection model in pathologies caused by pathogens

As previously mentioned, the use of the larvae of *G. mellonella* as an experimental model to analyze host-pathogen interactions as well as the efficacy of antimicrobial agents is well accepted by the entire scientific community worldwide.

Regarding the data, almost 3,000 scientific articles have been published in Medline Database on *G. mellonella* so far, of which over 1,000 were published from 2019 to 2022, therefore demonstrating the growing popularity of this infection model (**Figure 22**).

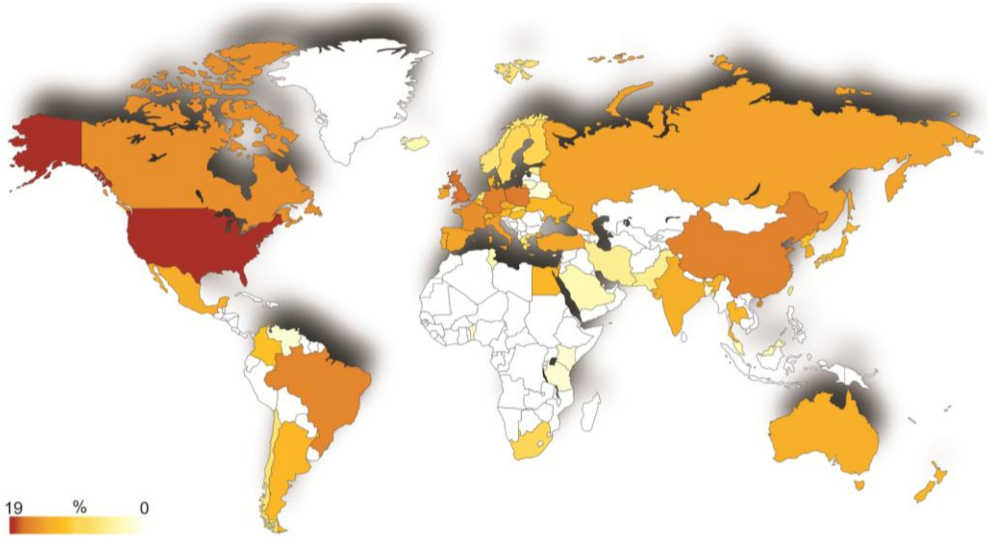


Figure 22. Locations and representativeness in percentage of worldwide institutions that perform studies with *G. mellonella*. Figure from (145).

A huge diversity of pathogens has been studied in *G. mellonella*–pathogen infection models leading to invaluable contributions to research with human and veterinary bacterial pathogens.

One of the most extensively studied parameter in *G. mellonella* is bacterial virulence, which is produced in these insects after a pathogenic inoculation. Most of *G. mellonella* studies with bacterial pathogens have been focused on Gram-negative bacteria (**Figure 23**) (145). The bacterial virulence can be evaluated in several ways such as analyzing the degree of melanization and the larval death over time (167), observing the morphology and measuring the density of hemocytes (168,169), estimating the expression levels of several enzymes and AMPs (170) and/or observing the histopathological effects produced in these worms (171).

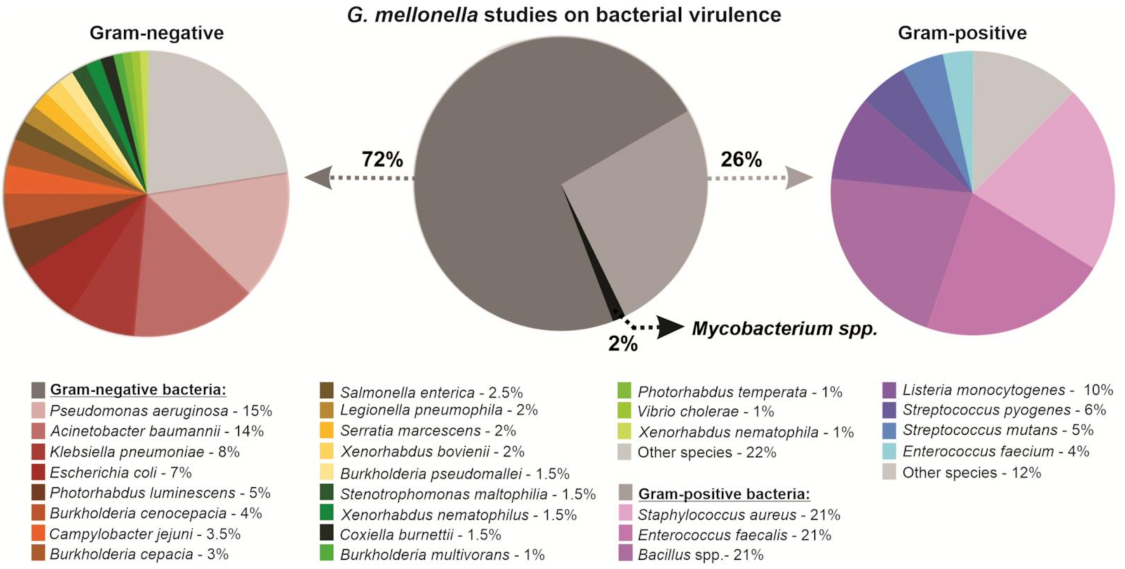


Figure 23. Use of *G. mellonella* model to study bacterial virulence. Figure from (145).

Apart from studying bacterial virulence, another hot topic studied in *G. mellonella* is the efficacy of several types of antimicrobial therapies against Gram-negative and Gram-positive bacteria (**Figure 24**) (145).

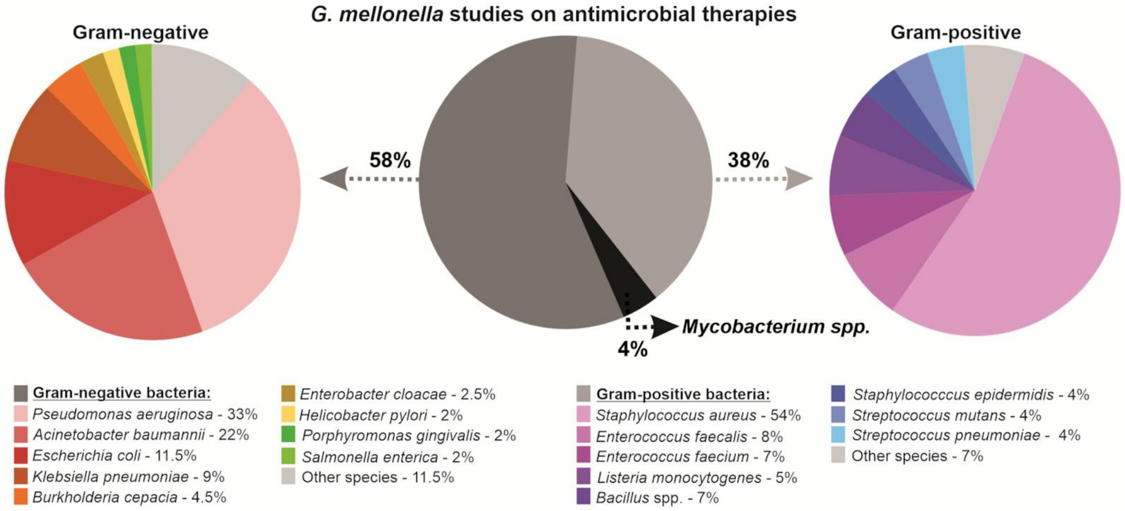


Figure 24. Use of *G. mellonella* model to evaluate antimicrobial therapies. Figure from (145).

From these studies, for an enormous variety of microorganisms, a positive correlation between virulence and host response has been stablished in both invertebrate and mammalian host models (164). Some examples of this kind of studies are shown in **Table S3, Annex III**.

Despite of the fact that there is a huge number of studies of *G. mellonella* as animal model for studying several pathogens, currently, there are no studies of this model with *B. fragilis*, so there is a need for the development of infection methods with these bacteria, and, thus, to see if this organism can be used as an infection model and a way of searching for new antimicrobial compounds for its treatment.

2. OBJECTIVES

To develop new effective therapeutic agents against BFT-3 protein, associated to different pathologies such as CRC, bacteremia, intestinal abscesses, diarrhea and intestinal inflammatory diseases.

To achieve this purpose, firstly it was required: 1/ To perform a functional characterization of the protein and study various properties of BFT-3, including the role of the zinc as cofactor. 2/ To identify compounds that interact with and inhibit BFT-3 through experimental molecular screening *in vitro*, followed by validation through *in vitro* techniques, cell cultures, and animal model assays.

The following general objectives have been stablished in Chapter I:

1. **To conduct a comprehensive protein characterization** to examine various aspects of BFT-3.
2. **To identify novel bioactive compounds** targeting the BFT-3 protein, which is relevant in the context of CRC.
3. **To investigate the effects of the selected compounds** on the BFT-3 protein using *in vitro* and *in vivo* systems.

Regarding the first objective, "*To conduct a comprehensive protein characterization to examine various aspects of BFT-3*", the following specific goals were proposed:

- To overexpress the BFT-3 protein in *E. coli* culture and to purify the recombinant protein by affinity chromatography.
- To examine the role of the zinc cofactor of BFT-3.

Regarding the second objective, "*To identify novel bioactive compounds targeting the BFT-3 protein, which is relevant in the context of CRC*", the following specific aims were proposed:

- To design and conduct a high-throughput screening (HTS) of compounds *in vitro* from a chemical library against BFT-3 as the target molecule. This aims to identify candidate compounds based on ligand-induced protein stabilization against thermal denaturation.
- To verify the interactions and determine the binding affinity between BFT-3 and the selected compounds in the screening.

Regarding the third objective, "*To investigate the effects of the selected compounds on the BFT-3 protein using in vitro and in vivo systems*", the following specific goals were proposed:

- To investigate the effects of the selected compounds *in vitro* on BFT-3 protease activity.

- To analyze the effects of the selected compounds *in vitro* using cancer cell lines exposed to BFT-3.
- To determine the residues implicated on BFT-3-compound interactions.
- To evaluate the antibacterial effects of the selected compounds on *B. fragilis* cultures.
- To implement and develop a novel animal model: *G. mellonella* as an infection animal model for *B. fragilis* infection.
- To analyze the antibacterial effects of the most promising compounds *in vivo* in *G. mellonella*.

3. MATERIALS AND METHODS

3.1. Plasmid construction

The proBFT-3 enterotoxin proprotein, i.e., without the signal peptide (residues A18-D397 detailed below in **Figure 25**; UniProt access code O86049), was cloned in the pET-28-based pCri-8a expression vector, which an attached fusion protein with a TEV-cleavable N-terminal His6-tag (**Figure 26 and Figure 27**). This plasmid was a kind of gift from our collaborator Dr. Francisco Xavier Gomis-Rüth from the Proteolysis Laboratory Group of the Molecular Biology Institute of Barcelona (IBMB).

```
ACSNEADSLTTSIDAPVTASIDLQSVSYTDLATQLNDVSDFGKMIILKDNGFNRQVHVSMDKRTKIQLDN
ENVRLFNGRDKDSTNFILGDEFAVLRFYRNGESISYIAYKEAQMMNEIAEFYAAPFKKTRAINEKEAFEC
IYDSRTRSAGKYPVSVKINVDKAKKILNLPECDYINDYIKTPQVPHGITESQTRAVPSEPKTVYVICLRE
NGSTVYPNEVSAQMQDAANSVYAVHGLKRYVNLHFVLYTTEYACPSGNADEGLDGFTASLKANPKAEGYD
DQIYFLIRWGTWDNNILGISWLNSYNVNTASDFKASGMSTTQLMYPGVMAHELGHILGANHADDPKDLMY
SKYTGYLFHLSEKNMDIIAKNLGWEIADGD
```

Figure 25. Cloned sequence in the vector.

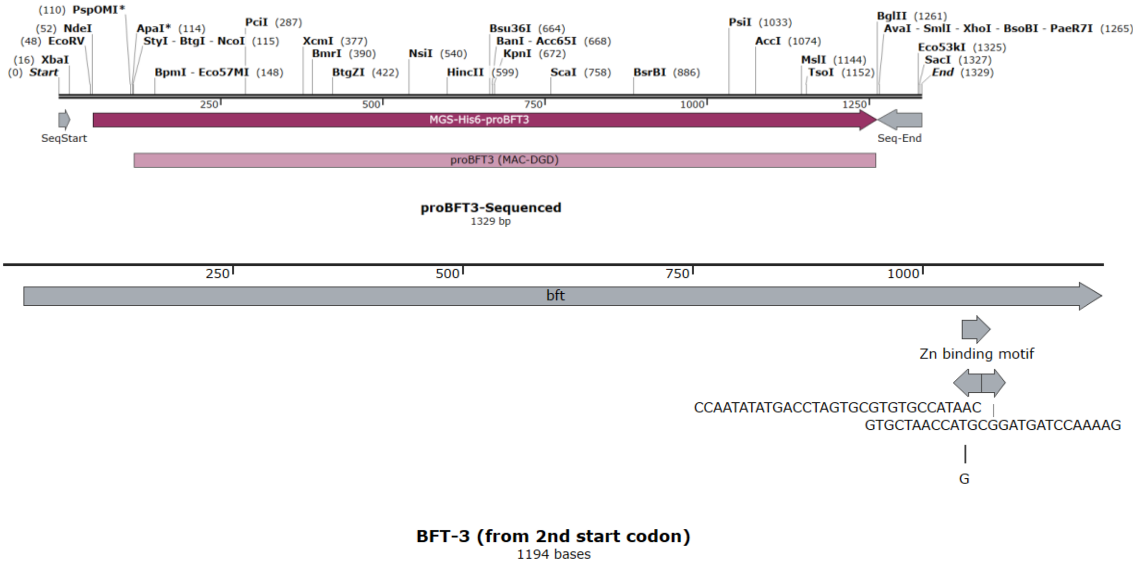

(110) PspOMI*
(52) NdeI
(48) EcoRV
(16) XbaI
(0) Start
ApaI* (114)
StyI - BtgI - NcoI (115)
BpmI - Eco57MI (148)
PciI (287)
XcmI (377)
BmrI (390)
BtgZI (422)
NsiI (540)
HincII (599)
Bsu36I (664)
BanI - Acc65I (668)
KpnI (672)
ScaI (758)
BsrBI (886)
PsiI (1033)
AccI (1074)
MslI (1144)
TsoI (1152)
BglII (1261)
AvaI - SmlI - XhoI - BsoBI - PaeR7I (1265)
Eco53kI (1325)
SacI (1327)
End (1329)
250
500
750
1000
1250
MGS-His6-proBFT3
SeqStart
Seq-End
proBFT3 (MAC-DGD)
proBFT3-Sequenced
1329 bp
250
500
750
1000
bft
Zn binding motif
CCAATATATGACCTAGTGCGTGTGCCATAAC
GTGCTAACCATGCGGATGATCCAAAAG
G
BFT-3 (from 2nd start codon)
1194 bases

Figure 26. Scheme of the proBFT-3 insert.

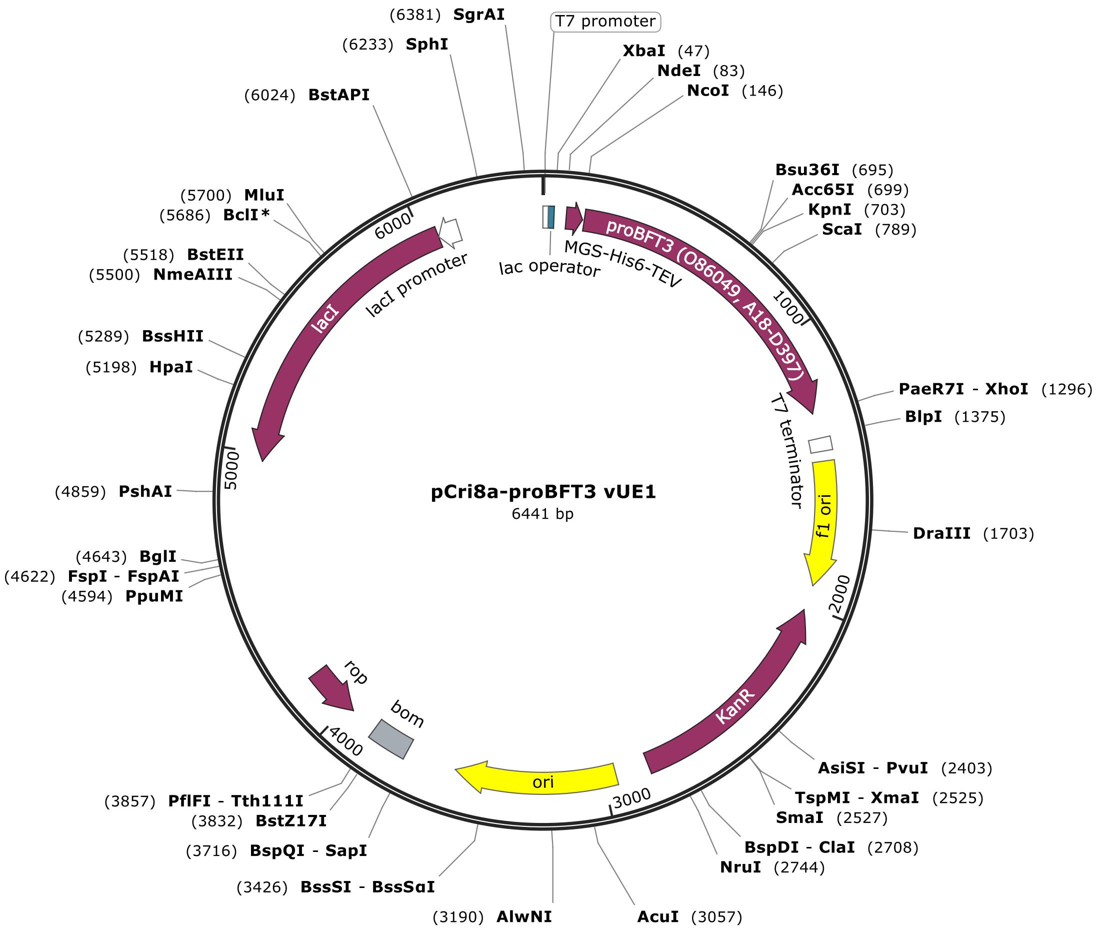

(6381) SgrAI
T7 promoter
(6233) SphI
XbaI (47)
NdeI (83)
NcoI (146)
(6024) BstAPI
Bsu36I (695)
Acc65I (699)
KpnI (703)
ScaI (789)
(5700) MluI
(5686) BclI*
(5518) BstEII
(5500) NmeAIII
(5289) BssHII
(5198) HpaI
6000
lacI promoter
lacI
lac operator
MGS-His6-TEV
proBFT3 (O86049, A18-D397)
1000
PaeR7I - XhoI (1296)
BlpI (1375)
T7 terminator
5000
(4859) PshAI
pCri8a-proBFT3 vUE1
6441 bp
f1 ori
DraIII (1703)
(4643) BglI
(4622) FspI - FspAI
(4594) PpuMI
2000
rop
bom
KanR
4000
ori
3000
AsiSI - PvuI (2403)
TspMI - XmaI (2525)
SmaI (2527)
BspDI - ClaI (2708)
NruI (2744)
(3857) PflFI - Tth111I
(3832) BstZ17I
(3716) BspQI - SapI
(3426) BssSI - BssSαI
(3190) AlwNI
AcuI (3057)

Figure 27. Plasmid construction based on pCri vector with the proBFT-3 insert.

3.2. Protein expression and purification

The proBFT-3 plasmid construct was transformed into the Star BL21 *E. coli* (DE3) strain which was grown in LB/Kanamycin (Kan) (50 µg/mL) at 37 °C overnight (see LB medium composition below in **Table 2**). The Kan antibiotic was purchased from Sigma-Aldrich. Positive colonies were those selected due to the Kan resistance acquired by plasmid incorporation. Negative control consisted of non-transformed cells spread in a LB/Kan plate.

Small-scale BL21 *E. coli* (DE3) cell cultures (250 mL) were grown in LB/Kan at 37 °C overnight (MaxQ5000 Orbital Shaker, from Barnstead Lab-Line). After that, they were used as pre-cultures (1:30) to inoculate large-scale 6-L cultures of LB/Kan (50 µg/mL), which were incubated under the same conditions (at 37 °C) until reaching an OD_{600} ≈ 0.6-0.7 (measured in a Ultrospec 6300 Pro Spectrophotometer, from Amersham Biosciences).

LB medium composition: g/L. Final pH 7.0	
Sodium chloride	5 g
Yeast extract	5 g
Tryptone	10 g

Table 2. LB medium composition.

Protein expression was induced with 1 mM isopropyl 1-thio-β-*D*-galactopyranoside (IPTG) (Sigma-Aldrich) at 18 °C overnight. Then, cells were harvested by centrifugation at 4 °C for 10 min at 10,000 g in an Avanti J-26 XP Centrifuge (Beckman Coulter).

After that, proBFT-3 transformed-cells were resuspended in 500 mM NaCl, 20 mM Tris-HCl, pH 7.4 as binding buffer. Cells were lysed by sonication (Vibra-Cell Ultrasonic Liquid Processor, from Sonics) on ice, after adding 20 U/mL benzonase (Merck-Millipore) to remove nucleic acids, 0.5 mg/mL lysozyme (Carbosynth), protease inhibitors cocktail (100 µM PMSF, from Sigma Aldrich; 1 mM Benzamidine, from Sigma Aldrich; and 5 µM Leupeptin, from Peptide Institute INC; being these ones the final concentrations after protease inhibitors cocktail addition to the cellular lysate). Cell debris was removed by 4-5 subsequent serial centrifugations at 20,000 g at 4 °C for 10 min.

Supernatants were clarified by subsequent filtration (0.45 µm-pore size membrane) to remove large impurities and then subjected to immobilized metal affinity chromatography (IMAC) purification in an ÄKTA FPLC (fast protein liquid chromatography) system (GE Healthcare Life Sciences). In order to obtain the desired protein, a cobalt HiTrap TALON column (GE Healthcare Life Sciences) was used, and an imidazole elution gradient (10-250 mM) was applied. Purification procedure monitoring and data storage was developed by the UNICORN control software (GE-Healthcare Life Sciences).

Ideally, only recombinant proteins would have been retained in the cobalt HiTrap TALON column when injecting the sample in the equilibrating binding buffer. Purity after each purification step was assessed by SDS-PAGE and pure protein fractions were pooled and dialyzed to remove imidazole in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4.

Final protein concentration was quantified by UV absorbance (NanoVue Spectrophotometer, from GE Healthcare Life Sciences) by using the theoretical extinction coefficient ε (proBFT-3: 54,780 M^{-1} cm^{-1}) at 280 nm. In each measurement, the UV absorbance at 350 nm was always considered as the background reference signal. DNA contamination in protein samples were always determined by calculating the ratio between the UV absorbance at 260 nm *vs.* the UV absorbance at 280 nm.

When needed, buffer exchange was performed using a 10 kDa-pore size ultrafiltration device (Amicon centrifugal filter, from Merck-Millipore) by centrifugation at 4,000 rpm and 4°C. Zinc-free proBFT-3 protein was obtained by adding ethylenediaminetetraacetic acid (EDTA, from Thermo Fisher Scientific) at a concentration not higher than 1 mM.

Whenever required, mature active BFT-3 was obtained from proBFT-3 by activation with trypsin (PAN-Biotech GmbH) added at a 1:100 mass ratio and incubated at room temperature for 3 h.

The correct identity of proteins was verified by mass spectrometry (MALDI-TOF/TOF) in the proteomics technical service at the Centro de Investigación Biomédica de Aragón (CIBA).

3.3. Circular dichroism

Circular dichroism (CD) is being increasingly recognized as a valuable technique for studying the structure of proteins in solution. CD is a spectroscopy technique based on the differential absorption of two circularly polarized components of equal intensity: one rotating counter-clockwise (left-handed) and the other clockwise (right-handed) circularly polarized light beams. CD signals only arise when differential absorption of radiation occurs, and thus spectral bands are easily assigned to distinct structural features of a molecule (172). Depending on the wavelength range, different structural and compositional aspects in proteins can be probed.

In the study of proteins by CD, the chromophores of interest consist of the peptide bond (absorption <240 nm), side chains of aromatic amino acids (absorption in the range 260-320 nm) and disulfide bonds (weak broad absorption bands at ~ 260 nm).

The information which can be obtained from CD experiments of proteins include: secondary structure composition (percentage of helix, sheet, turns, etc.) from the peptide bond region (240 nm and below), tertiary structure fingerprint (260-320 nm) from each aromatic amino acid wavelength profile, integrity of cofactor binding sites when they bound to their protein partner, conformational changes in proteins mainly caused by the binding of ligands or changes in environmental conditions, information of protein folding in the study of the refolding of denatured proteins, and data about the overall structure features of proteins to confirm the integrity of expressed domains of a multi-domain protein.

Employing a thermostated Chirascan spectrometer (Applied Photophysics) the CD spectra were recorded, using a quartz cuvette with a path length of 0.1 cm or 1 cm (Hellma Analytics), selecting a bandwidth of 1 nm, a spectral resolution of 0.5 nm and a response time of 5 s. Temperature was regulated by a Peltier unit and controlled using a temperature probe. Assays were performed in the far-UV range (190-260 nm, 0.1 cm path length) and in the near-UV range (250-310 nm, 1 cm path length), and 10 µM was the protein concentration. For optical components safety and scanning quality, oxygen was purged from the spectrometer through a nitrogen flow, 20 min before the measurements and during measurements.

The spectra obtained were baseline-corrected by subtracting the buffer signal and normalized by concentration and protein size in order to determine the mean residue ellipticity [ϴ] (mdeg cm^{-1} M^{-1}), according to the following equation [1].

$$[1] \qquad [\Theta] = \frac{\Theta}{10\, c\, l\, N};$$

where *ϴ* is the raw ellipticity measured (mdeg), *c* is the concentration of the protein (M), *l* is the optical path length (cm), and *N* is the number of amino acid residues.

3.4. Fluorescence spectroscopy

One of the most widely used techniques in protein structural studies is fluorescence spectroscopy (173). This technique takes advantage of the intrinsic characteristics of tryptophans (Trp), which are aromatic amino acids that absorb at 280 nm. The fluorescence exhibited by Trp is extremely dependent on its surrounding microenvironment, which makes possible to measure changes in protein structures by monitoring the fluorescence emission emitted by this amino acid. Specifically, Trp fluorescence undergoes an intensity reduction and shift to higher wavelengths (red-shift) when exposed to the solvent. Therefore, fluorescence spectroscopy is a good technique for monitoring the folding/unfolding of a protein (or even minor conformational changes) or for the study of protein interaction equilibrium. As mentioned above, this technique provides information on protein structures, but it has a clear limitation: there must be at least one Trp in the sequence of the protein under study (and, even a large number of Trp residues may be detrimental).

In addition, extrinsic fluorophores that bind to hydrophobic regions in proteins can be employed, so that they can be used to observe protein folding/unfolding under different environmental conditions. If under a given condition, the protein of interest shows an increase in fluorescence, it will indicate an initially buried region within the folded core structure becomes now exposed to the solvent, indicating that the protein under these conditions underwent a complete or partial unfolding event.

A Cary Eclipse spectrofluorometer (Agilent Technologies) interfaced with a thermostated multicell holder (Peltier unit) was employed to collect the fluorescence spectra. For both excitation (λ_{exc}) and emission (λ_{em}) wavelength filters, the slit widths were 5 nm. To assess the folding of proBFT-3, 2 µM solutions were used, and to evaluate the solvent-exposed molecular surface, the 8-anilino-1-naphthalenesulfonic acid (ANS) as extrinsic probe was used, with an excitation wavelength of 370 nm and recording the emission spectrum from 400 to 600 nm. Then, 2 µM proBFT-3 and 100 µM ANS were the fixed concentrations for the assays, which were carried out in a quartz cell with 1-cm path length (Hellma Analytics) at 25 °C. All measurements were performed in 150 mM NaCl, 20 mM Tris-HCl, pH 7.0, in absence or presence of 1 mM EDTA to remove the catalytic zinc from the protein when needed.

3.5. Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a readily accessible, rapid, and highly efficient biophysical technique that provides information about protein unfolding by monitoring the changes produced in fluorescence as a function of the temperature. This technique typically uses a hydrophobic fluorescent dye (i.e., an extrinsic probe, such as SYPRO Orange or ANS) that binds to proteins subject to study as they unfold (174).

By using DSF, the thermal stability of proBFT-3 and its modulation by zinc binding was evaluated in a Stratagene Mx3005P qPCR real-time thermal cycler (Agilent Technologies). The hydrophobic fluorescent dye used was the extrinsic fluorophore SYPRO Orange Protein Gel Stain (Thermo Fisher Scientific), which allowed recording the fluorescence emission intensity as a function of temperature in 150 mM NaCl, 20 mM Tris-HCl, pH 7.0, at a scanning rate of 1 °C/min using an excitation and an emission wavelength filter with nominal wavelengths λ_{exc} = 496 nm and λ_{em} = 610 nm, respectively. These filters were the closest available to the theoretical values of the fluorophore (λ_{exc} = 491 nm and λ_{em} = 586 nm). Briefly, along the protein unfolding process, SYPRO Orange binds to hydrophobic patches on the protein surface as they become exposed to the solvent and accessible to the extrinsic probe. As a result, the fluorescence intensity increases as the fluorescence quantum yield of the fluorophore is enhanced (175–177).

In order to analyze the role of zinc in the stabilization of the protein, 200 µM EDTA was added to the buffer to deplete the protein of zinc and, therefore, obtain the apozymogen or apoenzyme. Thermograms were obtained with 2 µM proBFT-3 and 5X SYPRO Orange. For zinc-bound proBFT-3, no exogenous zinc was added; therefore, zinc concentration was that corresponding to the endogenous zinc accompanying proBFT-3 along and after the purification.

A single unfolding transition model was considered for the experimental data analysis, providing relevant thermodynamic stability parameters such as the unfolding temperature, T_m, and the associated unfolding enthalpy, $\Delta H(T_m)$.

3.6. Differential scanning calorimetry

One of the most suitable techniques to study the energetics of protein folding-unfolding transitions is differential scanning calorimetry (DSC). By means of this technique, a thermodynamic characterization of the conformational changes induced by temperature changes in proteins can be achieved. In a DSC assay, the thermogram obtained is the result of continuously recording the apparent excess molar heat capacity of a protein solution as a function of temperature. By analyzing the thermogram, information on the relative heat capacity of a system as a function of temperature can be obtained, which is the fundamental information provided by a DSC assay (178), providing basic information about the structural stability parameters (unfolding temperature, T_m, unfolding enthalpy $\Delta H(T_m)$, and unfolding heat capacity, ΔC_P) of the protein and the internal cooperativity of the molecular structure.

Thermal stability of proBFT-3 was assessed by temperature unfolding transitions monitored by high-precision differential scanning calorimetry. The partial molar heat capacity of the protein in solution was measured as a function of temperature in an Auto-PEAQ-DSC (MicroCal, Malvern-Panalytical). Experiments were performed with a proBFT-3 protein solution at a concentration of 9-14 µM in 50 mM Tris-HCl, pH 8.0, and scanning from 15 to 95 °C at a rate of 60 °C/h. Different buffers were employed to obtain protein stability information under different conditions: with 2 mM/500 µM EDTA or 100/200 or 100/1000 uM $ZnCl_2$/$ZnSO_4$ or 150 mM NaCl. Reference solutions containing buffers, but lacking proBFT-3, were treated similarly to record the baseline of the instrument before experiments. No protein precipitation/aggregation occurred during thermal denaturations. For the data analysis, thermograms were baseline-corrected and analyzed using the software Origin 7 (OriginLab).

3.7. Chemical libraries

One commercially available chemical library was employed for the experimental molecular screening: Prestwick Chemical Library (Prestwick Chemical) with 1,120 FDA-approved drugs, providing a large chemical and pharmacological diversity and most of them complying with Lipinski rules, very often employed to guarantee oral bioavailability (179). Compounds were dissolved in 100% dimethyl sulfoxide (DMSO) at a final concentration of 4 mM, and they were arranged in 96-well plates preserved at -20 °C while not in use. ADMET (chemical absorption, distribution, metabolism, excretion, and toxicity) properties, therapeutic indications as well as toxicity and safety in humans are available from the manufacturer for Prestwick compounds.

3.8. Experimental ligand screening: thermal shift assays

A high throughput screening (HTS) procedure, based on the thermal-shift assay (TSA, ligand-induced protein stabilization against thermal denaturation) by DSF, was employed to identify potential ligands for proBFT-3 from the Prestwick Chemical.

This procedure was previously employed in our research group for identifying small-molecule compounds acting as inhibitors of NUPR1 (target protein in Chapter II) (180), inhibitors of NS3 protease from hepatitis C virus (181), inhibitors of *Helicobacter pylori* flavodoxin (182) and pharmacological chaperones of human phenylalanine hydroxylase (183).

By using TSA, in which the thermal denaturation of proteins was monitored using fluorescent amphipathic dyes, the stability of a protein can be easily determined by fluorescence detection in a real-time PCR instrument. In this way, a screening of compounds in 96-well format was performed, where the thermal shift identified the conditions under which the stability of a recombinant protein was enhanced by the presence of a given compound. Thus, it was possible to analyze protein-ligand interactions in which the compounds used as decoys increased the stability of the protein (184).

As proBFT-3 is a zinc-dependent protein, a screening procedure tailored to the identification of potential ligands of the zinc-free conformational state was employed following the same strategy previously carried out in our research group in the identification of inhibitors for the NS3 protease from hepatitis C virus (181).

The stabilizing compounds from the Prestwick Library were selected by using a FluoDia T70 fluorescence microplate reader (Photon Technology International). By this methodology, it was possible to detect ligand-induced protein stabilization against thermal denaturation (175–177,185) by monitoring the thermal denaturation of proBFT-3 in presence of the extrinsic fluorescent probe SYPRO Orange.

Specifically, 1 µM proBFT-3, 100 µM compound, and 1X SYPRO Orange in 150 mM NaCl, 5 mM EDTA, 20 mM Tris-HCl, pH 7.4, in a final volume of 100 µL, were dispensed into 96-well microplates (ThermoFast 96 skirted plates, from Thermo Scientific). Therefore, each reaction volume had 2.5% DMSO from the tester stock solutions. Besides, proBFT-3 with 2.5% DMSO as control samples were routinely included in each microplate. To prevent the evaporation of the samples 20 µL mineral oil were employed to overlay the solutions and then incubated at 25 °C for 30 min. Thermal unfolding fluorescence traces were obtained from 25 to 75 °C, at a 1 °C/min scan rate. The λ_{exc} and λ_{em} filters employed (λ_{exc} = 470 and λ_{em} = 570 nm) the closest available to the theoretical values of the fluorophore (λ_{exc} = 491 and λ_{em} = 586 nm). Thermal equilibrium was reached at each temperature by considering an equilibration time of 1 min before each measurement, which corresponded to an approximate operational heating rate of ~0.25 °C/min. Control experiments in the absence of EDTA and/or ligands were routinely performed in each microplate.

The midpoint temperature (T_m) can be operationally defined as the temperature for maximal slope in the thermal unfolding curve (185). Those compounds able to increase the T_m by at least 3 °C, compared to the internal controls in each microplate (8 experiments without added ligand and performed in the same microplate), were identified as hits.

3.9. Nomenclature of selected compounds in the *in vitro* experimental ligand screening

The chemical compounds selected in the experimental screening by TSA in Chapter I were internally coded with a name for a quick and easy identification.

The selected compounds from the Prestwick Chemical Library were named MOAX, with "X" being the compound identifier number from 1 to 11 (i.e., MOA1, MOA2, MOA3…).

3.10. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is one of the most widely used analytical techniques in the study of interactions of biological interest, such as intermolecular and intramolecular, protein-ligand, protein-protein, and protein-DNA/RNA associations (186). The basis of this technique consists of directly measuring the heat released (exothermic interactions) or absorbed (endothermic interactions) in a reaction brought into play as a result of the previously mentioned interactions (187). The success of this technique is due to its instrumental and experimental simplicity together with the fact that, in principle, any interaction releases or absorbs heat. Since this heat is measured at constant pressure, what is being measured is the true enthalpy of the underlying reaction. Thus, if the net thermal effect of the interaction is sufficiently large, the stoichiometry, enthalpy of formation (ΔH) and equilibrium interaction constants (association constant, K_a; dissociation constant, $K_{d.}$) of these complexes can be determined by this technique [2]. Consequently, the Gibbs energy of formation of the complex (ΔG) can be calculated [3], and subsequently the entropy (ΔS) of the process [4]. Therefore, a complete characterization of the interaction can be achieved for many biological interactions (186,188).

$$[2] \qquad Kd = \frac{1}{Ka}$$

$$[3] \qquad \Delta G = -RT\, lnKa$$

$$[4] \qquad \Delta S = \frac{\Delta H - \Delta G}{T}$$

where R is the molar gas constant (8.314 J/K·mol) and T is the absolute temperature (K).

Calorimeters are highly sophisticated equipment used in this technique to determine the amount of heat gained or released by the system when two molecules interact. The basic scheme of a standard ITC instrument consists on the compensation of the thermal effect in an adiabatic environment. In this type of calorimeters there are two identical cells, the reference and the sample cells, and, during an experiment, a small constant power triggers the dynamic compensation system which equalizes the temperatures of both cells at any moment and produces the experimental baseline. An exothermic or endothermic effect, produced by ligand addition in the sample cell from a controlled syringe, transiently changes the temperature in the sample cell and the feedback actuator of the compensation system provides an additional thermal power (by excess or by defect) to the sample cell, which is the direct measured signal and resulting in a characteristic deflection peak from the baseline. When all the heat associated with the ligand addition is dissipated, the thermal power returns to the initial value (baseline). The signal associated with the transient thermal effect of a reaction presents a peak with the characteristic shape of the response of a second order system to a short duration power pulse. Thus, the sequence of ligand injections is reflected in a sequence of signal peaks, called thermogram (thermal power *vs.* time). From the area delimited by the peak and the baseline interpolated between the stationary signal before and after the injection, the corresponding heat effect per injection is obtained (189,190). The binding or interaction isotherm is constructed by plotting the heat effect per injection as a function of the molar ratio (quotient between the total concentration of ligand in cell and the total concentration of protein in cell). Analyzing the binding isotherm employing the appropriate binding model and non-linear least-squares regression data analysis, the binding parameters for a given interaction can be estimated.

All instrument-dependent experimental variable parameters, such as temperature, number of injections, injection volume, stirrer rotation speed, time between injections, etc., are programmed from the instrument software, so that the experiment is performed fully automatically.

The interaction of selected compounds with proBFT-3 and BFT-3 (activated with trypsin) was evaluated with a high-sensitivity isothermal titration calorimeter Auto-iTC200 (MicroCal, Malvern-Panalytical). To avoid problems related to bubble formation and to ensure the good quality of the assays, both protein samples and reference solutions were properly degassed and carefully loaded into the calorimeter cells. Experiments were performed in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4 at 25 °C. In the case of experiments with zinc-free protein, 1 mM EDTA was added to the protein solution. 20 µM of the protein solution was titrated in the calorimetric cell with 300 µM of the compound solution through a series of 19 injections of 2 µL (the stirring speed was 750 rpm and the reference power was 10 µcal/s). After integrating the calorimetric signal in the thermogram, the heat effects after each ligand injection were obtained. The difference between the reaction heat and the corresponding heat of dilution provided the heat due to the binding reaction. The heat of dilution was estimated as a constant value throughout the experiment, therefore included as an adjustable parameter in the analysis. In order to account for potential unspecific phenomena, i.e., solution composition mismatches or self-association of compounds, compound injected into buffers as control experiments were performed under the same experimental conditions. Non-linear least-squares regression data analysis of the experimental data, applying a model considering a single ligand binding site in the protein, was employed to obtain the ΔH and the K_a of the binding reaction. Uncertainties for the estimated parameters were estimated according to asymmetric profile likelihood 95% confidence intervals (191). All the experiments were carried out in replicates and data were analyzed using in-house developed software implemented in Origin 7 (OriginLab, Northampton, MA).

3.11. *In vitro* enzymatic inhibition assay

In order to determine the potential of the selected compounds to inhibit the enzymatic activity of proteins, *in vitro* enzyme inhibition assays were needed.

To analyze the *in vitro* activity of BFT-3 and the inhibitory effect of the selected compounds, proteolytic assays were performed using BODIPY casein as the substrate (EnzChek Protease Assay Kit, from Thermo Fisher Scientific) according to the manufacturer's instructions.

Molecular Probes' EnzChek Protease Assay Kit is a fast, simple and direct fluorescence-based assay for detecting metallo-, serine, acid, and sulfhydryl proteases. This kit contains casein derivatives that are heavily labelled with a pH-insensitive green fluorescent BODIPY® FL dye, resulting in almost total quenching of the fluorescence of the conjugate. Protease-catalyzed hydrolysis releases highly fluorescent BODIPY FL dye-labelled peptides (**Figure 28**). The accompanying increase in fluorescence, which can be measured with a microplate reader, is proportional to protease activity (192).

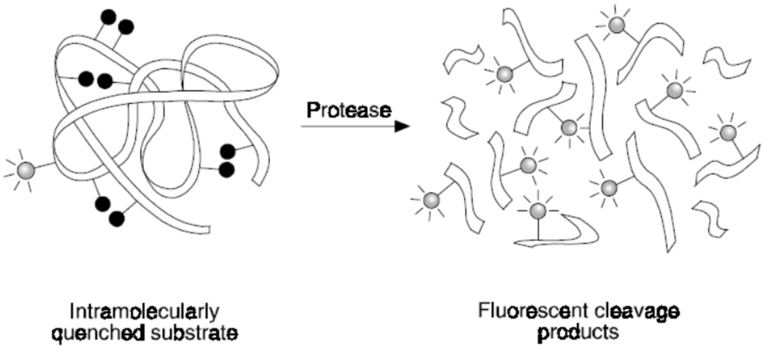

Protease
Intramolecularly quenched substrate
Fluorescent cleavage products

Figure 28. Principle of protease activity detection employed in the EnzChek Protease Assay Kit.

First, 3 µM proBFT-3 in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4 were activated by adding trypsin at a 1:100 molar ratio in a final volume of 50 µL, and then incubated for 3 h at room temperature. After that, two-fold serial dilutions were made with each compound from an initial concentration of 3 mM in 7% DMSO, 0.1 mM NaN_3, 10 mM Tris-HCl, pH 7.8, and then 50 µL of each dilution were added to the protein sample and subsequently incubated for 1-2 h. After that, 50 µL BODIPY substrate were added to each protein-compound sample and the fluorescence intensity was continuously recorded (λ_{exc} = 492 and λ_{em} = 516 nm, the available wavelengths closer to the recommended 505/513 nm for the substrate) at 37 °C for 1 h in a Stratagene Mx3005P real-time qPCR instrument. Control experiments (proBFT-3 without trypsin activation, trypsin, and buffer alone) were carried out under the same experimental conditions.

3.12. Crystallographic assays

The protein crystallographic assays were performed in collaboration with Dr. Xavier Gomis-Rüth and Dr. Ulrich Eckhard of the Institute of Molecular Biology of Barcelona (IBMB).

One of the main techniques for the determination of protein structures is X-ray crystallography. Around 85% of the protein structures so far described have been determined using X-ray crystallography. Knowing the 3D structures of proteins is compulsory for a large variety of applications, such as biotechnology, drug design, biomedicine, and basic research, as well as a validation tool for protein modifications, structural accuracy and ligand binding (193).

Employing Vivaspin 20 centrifugal concentrators with a molecular-mass cutoff of 10-kDa (Sartorius, PES) the proBFT-3 protein was concentrated to 10-15 mg/mL and then polished with 50 mM NaCl, 20 mM Tris-HCl, pH 8.0, as running buffer, using a Superdex S200 gel filtration column which was connected to an ÄKTA Purifier 10 (GE Healthcare Life Sciences). The peak fractions obtained were pooled and re-concentrated to ~10 mg/mL, quickly frozen in liquid nitrogen in aliquots of 50 µL and stored until further use at -80 °C. Using a Phoenix dispenser robot (Art Robbins Instruments) in sitting-drop vapor-diffusion setup, high-throughput crystallization screenings were performed by mixing 100 nL of proBFT-3 with 100 nL of reservoir solution. Commercial crystallization screens including JCSG+, PACT Premier, BCS, and LFS (all from Molecular Dimensions), and PEGRX, SaltRX, and Index Screen (Hampton Research) were used in duplicate setups to allow for incubation at 20 and 4 °C in Bruker AXS Crystal Farms. Initial crystallization hits were subsequently optimized by hand and upscaled to 0.5 µL drops utilizing standard 2D grid screening in sitting-drop vapor diffusion format.

Suitable crystals of unbound proBFT-3 in a new orthorhombic space group (NATI_ORTH) and in complexes with MOA4, MOA9 and MOA10 inhibitors in orthorhombic (ORTH) or tetragonal (TETR) space groups were obtained at 4 °C from 0.5 µL drops of proBFT-3 solution (at 7.9 mg/mL in 50 mM sodium chloride, 20 mM Tris-HCl, pH 8.0, supplemented with 5 µM $ZnCl_2$ and 400 µM of the respective inhibitor) and 0.5 µL of reservoir solution. The final crystallization conditions were 18% PEG 3350, 0.2 M MgHCOO, pH 5.9 for NATI_ORTH; 20% PEG 3350, 0.2 M $(NH_4)_2SO_4$ for the MOA4 TETR complex; 14% PEG 3350, 0.2 M $MgCl_2$, and 25% PEG 3350, 0.1 M Bis-Tris, pH 5.5, 0.2 M NH_4CH_3COO, for the MOA9 TETR and ORTH complexes, respectively; and 22% PEG 3350, 0.1 M Bis-Tris-HCl, pH 5.5, 0.2 M $(NH_4)_2SO_4$ for the MOA10 TETR complex. All crystals typically appeared within 7 days, contained two molecules per asymmetric unit, and were harvested and cryoprotected by soaking for 15-30 s in mother liquor supplemented with 2.5 M L-proline and 1 mM of the respective inhibitor. Of note, all inhibitor stocks were originally at 20 mM in 100% DMSO. Therefore, crystallization conditions and cryoprotection buffers contained 2% and 5% DMSO, respectively, in all cases.

3.13. X-ray crystallographic assays: structure determination and analysis

Datasets of the complete diffractions of both unbound and inhibitor-bound proBFT-3 crystals were achieved at i04-1 and XALOC beamlines of the Diamond and ALBA synchrotrons, respectively. Data were integrated, scaled, merged and reduced using XDS (194) and XSCALE (195), and transformed with XDSCONV to the reflection file format MTZ for use by the PHENIX (196), BUSTER/TNT (197), and CCP4 (198) program packages. Structures were solved by molecular replacement using PHASER (199) with one protomer of the PDB entry 3P24 (67) as a search model. Data collection and processing statistics are listed in **Table S4, Annex III**. Protein-ligand interactions were evaluated using the PLIP web tool with default settings (200,201), and both protein-protein and protein-ligand interaction areas were calculated using the PDBePISA webserver (202). Molecular graphics representations were created using an open-source build of PYMOL version 2.5 Plus: The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.) (203).

3.14. Cell culture assays

Cell culture assays were performed by using HT-29 (human colon adenocarcinoma) and HeLa (human cervix epithelioid carcinoma) cells obtained from ATCC and maintained in Dulbecco's odified Eagle's Medium (DMEM; PAN-Biotech GmbH) supplemented with 10% FBS (Fetal Bovine Serum) (PAN-Biotech GmbH, Aidenbach, Germany), 1% penicillin/streptomycin (Thermo Fisher Scientific), and 1% NEAAs (non-essential amino acids) (Thermo Fisher Scientific) with 5% CO_2 at 37 °C.

3.15. Cytotoxicity assays

In vitro cytotoxicity testing provides a crucial means of ranking compounds for consideration in drug discovery due to the fact that testing the viability effects of compounds on cell cultures could be used as a predictor of potential toxic effects in whole animals. In this way, it was possible to identify whether potential candidates selected in the screening have undesirable side effects in cell cultures, thus not being subjects of larger scale trials with the ultimate goal of clinical trials in humans. In addition, cytotoxicity data are required to design and plan compound efficacy assays by providing the practical concentration range appropriate for each compound.

Cellular cytotoxicity of the selected compounds was assessed in two different cell lines: HT-29 and HeLa cells. Cells were plated and incubated in 96-well plates (9,000 cells/100 µL/well for HT-29 cells and 8,000 cells/100 µL/well for HeLa cells) with FBS-supplemented DMEM without phenol red (PAN-Biotech GmbH) for 48 h. Two-fold serial dilutions of compounds were added to the cells up to 400 µM as the maximal concentration, and allowed to proliferate for other 48 h. After this period, the cytotoxicity produced by the compounds, measured as a cell viability parameter (percentage, taken the untreated control cells as a reference), was evaluated with the CellTiter 96 AQueous One Solution Cell Proliferation Assay Kit (Promega Corporation). For this measurement, 20 µL of CellTiter buffer was 1:4 diluted in FBS-supplemented DMEM without phenol red, and then added to each well with the samples in the 96-well assay plates. The plates were incubated for 2 h in a humidified atmosphere at 37 °C with 5% CO_2. Finally, the absorbance was registered at 490 nm, with a background correction at 800 nm (in order to minimize experimental uncertainties in the measurement), and the readout was directly proportional to the number of alive cells in the culture. The 50% cytotoxic concentration (CC_{50}), or when appropriate the half maximal inhibitory concentration (IC_{50}), was defined as the compound concentration that reduced the cell metabolism activity by 50%. Each condition was performed in triplicate and all experiments were repeated at least two times. Mean values and standard deviations (SD) were calculated and represented in the cytotoxicity.

3.16. Microscopy assays

BFT-3 is capable of producing morphological changes in cell cultures easily observable by optical microscopy at very low toxin concentrations, as previously described in the introduction of Chapter I. For this reason, it was verified in cell cultures of HT-29 cells whether the protein we were working with produced the same effects already known.

HT-29 cells were plated and incubated in 96-well plates (9,000 cells/100 µL/well) with FBS-supplemented DMEM without phenol red in a humidified atmosphere at 37 °C with 5% CO_2 for 48 h. Different concentrations (0, 225 nM or 2 µM) of proBFT-3 in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4 were added to cells and allowed them to proliferate for 48 h. After this incubation time, the morphological changes produced by the proBFT-3 were evaluated by taking images in bright field under a confocal microscope (DMI 6000B, Leica). Each experiment was repeated at least three times.

It was also tested whether the compounds were able to inhibit the morphological changes produced by proBFT-3 in cells. For this purpose, the above protocol was carried out, but with the additional step of adding two-fold serial dilutions of MOA compounds up to 120 µM as the maximal concentration at the same time step at which proBFT-3 was added (in this case the concentration for proBFT-3 was 775 nM).

3.17. Western blot assays

As explained in the introduction section, one of the cellular effects of BFT-3 is the cleavage of E-cadherin in cells. In order to evaluate the inhibitory capacity of selected compounds on E-cadherin processing by proBFT-3, western-blot assays were carried out to detect the hydrolysis of E-cadherin in HT-29 cell samples.

HT-29 cells were seeded in 24-well plates (90,000 cells/500 µL/well) with FBS-supplemented DMEM without phenol red. After 48 h, proBFT-3 was incubated with or without selected compounds in phosphate-buffered saline (PBS) for 3 h. Then, DMEM without phenol red was added to the samples achieving a 25 nM proBFT-3 final concentration. Cell media from each well was removed and the proBFT-3 or proBFT-3/compound solutions were added to the plates. Of note, in the absence of inhibitor, proBFT-3 is expected to be either activated autoproteolytically or by a secreted protease from the cells (i.e., proBFT-3 was not *in vitro* activated with trypsin). After 24 h, cell lysates were obtained using cold radioimmunoprecipitation assay buffer (RIPA), which consisted of 150 mM NaCl, 0.1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl, pH 8.0, and protease inhibitors. To obtain the protein extracts, samples were incubated on ice for 15 min and then diluted into 4X Laemmli loading buffer (Bio-Rad). Then, samples were heated to 95 °C for 5 min and subjected to 10% SDS-PAGE in a Mini-PROTEAN system (Bio-Rad) operated at 140 V. Using Mini-PROTEAN adaptors under wet and cold conditions at 400 mA for 60 min, the proteins were transferred to a PVDF membrane. 5% non-fat dry milk in Tris-buffered saline (TBS) was used to block the membranes at room temperature for 1 h and then incubated at 4 °C overnight with the respective primary antibody (E-cadherin Antibody or GAPDH Antibody, **Table 3**). The antibodies were previously diluted 1:2,500 with blocking solution supplemented with 0.1% Tween 20 (TBS-T). After three washes with TBS-T, membranes were incubated with the secondary antibody (previously diluted 1:10,000 with TBS-T), which was marked with Alexa Fluor Plus 647 fluorescence label (**Table 3**), at room temperature during 1 h. After three further washes with TBS-T, the signal was recorded in a ChemiDoc Gel Imaging System (Bio-Rad).

Primary antibody	Description	Host species	Species reactivity	Dilution	Reference
Anti-E-cadherin	E-cadherin Polyclonal Antibody (IgG)	Rabbit	Human	1:2,500	PA5-32178 Thermo Fisher Scientific
Anti-GAPDH	GAPDH Polyclonal Antibody (IgG) - Loading Control	Rabbit	Human	1:2,500	ab9485 Abcam
Secondary antibody	**Description**	**Host species**	**Species reactivity**	**Dilution**	**Reference**
Alexa Fluor Plus 647	Anti-Rabbit IgG (H+L) Highly Cross-Absorbed Secondary Antibody	Goat	Rabbit	1:10,000	A32733 Thermo Fisher Scientific

Table 3. Primary and secondary antibodies employed in western blot assays.

3.18. *B. fragilis* strains

The following strains of *B. fragilis* were supplied by ATCC. Strains have been purchased in collaboration with Dr. Jose Antonio Ainsa and Dr. Santiago Ramon's research group from the Microbiology Department of the University of Zaragoza. Two *B. fragilis* strains have been employed: Enterotoxigenic *B. fragilis* (ETBF) (ATCC 43858), which is capable of producing the BFT-3 toxin, and Non-enterotoxigenic *B. fragilis* (NTBF) (EUCAST clinical strain: 617161), which does not produce the toxin.

3.19. *B. fragilis* cultures

Bacterial stocks were preserved frozen in solution or with cryo-beads at -80 °C. To work with *B. fragilis* strains, since they are anaerobic, special conditions were required (204,205). The anaerobic glove box has become the method of choice for growing anaerobes in most research laboratories. However, it is also possible to use anaerobic jars (206).

In this way, anaerobic chambers and sachets (GasPak EZ Anaerobe Container System, 260678, from BD) were used, which are capable of removing the oxygen inside the chamber. These multi-use systems produce atmospheres suitable to support the cultivation of anaerobic, microaerophilic, or capnophilic bacteria by use of gas generating sachets inside multi-use incubation containers. This kind of sachets consist of a reagent sachet containing inorganic carbonate, activated carbon, ascorbic acid and water. Briefly, the sachet becomes activated by exposure to air when it is removed from the outer wrapper. Once active, it should be placed in the incubation container, which must be sealed quickly. Then, the sachet rapidly reduces the oxygen concentration within the container while the inorganic carbonate produces carbon dioxide. These systems produce an anaerobic atmosphere within 2.5 h with less than 1% oxygen, and greater than or equal to 13% carbon dioxide within 24 h, leading to ideal conditions to the cultivation of anaerobic bacteria (207).

The medium used for *B. fragilis* cultivation was *Brucella* broth (see composition in **Table 4**) supplemented with 5% defibrinated sheep blood (Thermo Scientific), 5 µg/mL hemin (Sigma-Aldrich) diluted from a 10 mg/mL stock dissolved into DMSO and syringe-filtered (Acrodisc 13 mm, 0.2 µm pore size, HT Tuffryn Membrane, PN4454, from Pall Corporation) and 1 µg/mL vitamin K1 (Sigma-Aldrich) diluted from a 5 mg/mL stock dissolved in acetone and syringe-filtered (204).

***Brucella* broth composition**: g/L. Final pH 7.0	
Casein digested by pancreatic enzymes	10.0 g
Animal tissue digested by peptic enzymes	10.0 g
Dextrose	1.0 g
Yeast extract	2.0 g
Sodium chloride	5.0 g
Sodium bisulfite	0.1 g

Table 4. *Brucella* broth composition.

Briefly, routine maintenance of *B. fragilis* cultures was performed in liquid medium every 48 to 72 h, using autoclaved tubes with *Brucella* broth supplemented as described above. Depending on the experiment to be carried out, the bacterial preparations were made in liquid medium as explained above, or it was necessary to generate *B. fragilis* cultures in solid medium. For this purpose, sterile Petri dishes were used, to which *Brucella* broth supplemented with the aforementioned was added and the presence of 1.5% agar (European Bacteriological Agar, from Condalab). In addition, for the MIC_{90}, synergies and TKA determination assays (*see materials and methods explained below*), defibrinated sheep blood was replaced by FBS in the same proportion. This change was necessary when performing the aforementioned assays, since the optical density of the bacterial cultures must be measured in order to determine the number of bacteria required for each experiment, and the presence of blood in the culture medium causes it to become dark, thus making it impossible to obtain accurate values of the optical density.

Both the maintenance as well as all experiments carried out with *B. fragilis* cultures were performed following the recommendations stablished for Class II Biological Safety Cabinets.

3.20. Determination of the antibacterial activity of the compounds: liquid minimum inhibitory concentration assays (liquid MIC_{90} assays)

The MIC (minimum inhibitory concentration) parameter is widely considered a common measure and primary readout to assess antimicrobial activities (208). MIC_{90} is defined as the minimum inhibitory concentration of an antimicrobial/compound at which 90% of the isolates are inhibited.

The antimicrobial activity of HTS hits were tested against both bacterial strains. Concretely, the cytotoxicity of the top compounds as well as that of some antibiotics used clinically for the treatment of *B. fragilis* infections was evaluated. The selected clinical antibiotics were Metronidazole CRS (MTZ; European Pharmacopoeia Reference Standard) and Cefoxitin sodium CRS (FOX; European Pharmacopoeia Reference Standard).

Briefly, compounds at 2X (maximal compound concentration was 20 mM) were diluted in *Brucella* broth medium supplemented as previously described, and then 2X compounds were added (150 µL/well of each mix) to row A in 96-well plates (flat bottom, Tissue Culture Test Plate 96F) per triplicate. After that, 75 µL/well of supplemented *Brucella* broth medium with the corresponding DMSO percentage, depending on each assay, were added to rows B-H in the 96-well plates. Then, 2-fold serial dilutions were made by transferring 75 µL from row A to row B, mixed and continued until row G (75 µL taken from row G were discarded). Positive controls without any hit compound (75 µL of medium + 75 µL of inoculum) and negative controls without any hit compound or inoculum (150 µL of medium with the appropriate DMSO percentage) were placed in row H.

Then, optimal optical densities (10^6 cell/mL) for 2X inoculums of each strain were prepared (assuming $OD_{600nm} = 0.88$ corresponds to $4 \cdot 10^8$ cell/mL, in a Ultrospec 10 Cell Density Meter, from Biochrom) with fresh supplemented medium (10 mL were needed per plate). After preparing the bacterial inoculums at the desired concentration, 75 µL/well of 2X inoculums were added to all 96-well plates (excluding negative control wells), and then 1:2 dilutions of the 2X inoculums were obtained, resulting in this way the 1X inoculums ($5 \cdot 10^5$ cell/mL). Plates were incubated in a bacterial incubator to allow them to grow at 37 °C for 48 h. After this period, the protocol for measuring cell viability was performed (*see below*).

3.21. Liquid media HTSS (high throughput synergy screening): synergy assays

Normally, when two compounds act simultaneously, the combination effect of these interactions can be classified into three main types (209):

- Additive/non-interactive: if the effect of the drug combination is similar to the addition of the effects of the two substances acting individually meaning that it is a pure summation effect.
- Synergistic/potentiation: if the combined effect of the constituents is larger than the additive effect.
- Antagonistic: if the combined effect results in a less than the additive effect.

Briefly, synergy can occur through a variety of mechanisms (210), which include:

- Pharmacodynamic synergism through multi-target effects.
- Pharmacokinetic synergism through modulation of drug transport, permeation and bioavailability.
- Elimination of adverse effects.
- Targeting disease resistance mechanisms.

Therefore, synergy between two (or more) drugs occurs when their activities within the combination are improved over the sum of their individual separate effects (211).

In these assays, a primary compound (PC) was established as the compound against which all other compounds, called secondary compounds (SC), were to be evaluated. For each experiment, because we are studying two strains of *B. fragilis*, six 96-well flat bottom plates were needed for testing different working conditions (plate 1: nontoxigenic strain/no PC with SC; plate 2: nontoxigenic strain/ 1/2X MIC_{90} PC with SC; plate 3: nontoxigenic strain/ 1/4X MIC_{90} PC with SC; plate 4: enterotoxigenic strain/ no PC with SC; plate 5: enterotoxigenic strain/ 1/2X MIC_{90} PC with SC; plate 6: enterotoxigenic strain/ 1/4X MIC_{90} PC with SC). Therefore, in order to evaluate synergistic effects of compounds in *B. fragilis*, assays were performed employing some clinical antimicrobials as PC (MTZ and FOX) and the selected MOA4 compound as SC. Different working conditions were needed in six 96-well flat bottom plates: 1) nontoxigenic/enterotoxigenic strain and no PC with MOA4; 2) nontoxigenic/enterotoxigenic strain and 1/2X MIC_{90} MTZ/FOX with MOA4; 3) nontoxigenic/enterotoxigenic strain and 1/4X MIC_{90} MTZ/FOX with MOA4; 4) nontoxigenic/enterotoxigenic strain and 1/8X MIC_{90} MTZ/FOX with MOA4. In all cases MOA4 was tested at different serial dilution concentrations.

The stock solutions, including PC and SC, were prepared at the required concentration in DMSO, taking into account that the 2X compound concentration of PC/SC should have the same proportion of DMSO. The 2X SC dilution (final volume/plate due to duplicates: 150 µL) were prepared in *Brucella* broth medium supplemented as previously described and added to the row A in each working plate. Then, in those working plates, 75 µL of supplemented medium with the same percentage of DMSO were added to rows B to G. For each compound, 2-fold serial dilutions were prepared by transferring 75 µL/well from rows A to G in the same column (discarding the 75 µL from rows G). The following step consisted on preparing mixes with medium, 2X inoculums for each strain (1X: $2 \cdot 10^5$ cell/well, assuming OD_{600nm}=0,88 is $4 \cdot 10^8$ cell/mL) and PC at 2X the subinhibitory MIC_{90} required: 1/2X MIC_{90}, 1/4X MIC_{90} (10 mL as final volume/plate). 75 µL/well of the prepared mixes were added to the working plates. Positive and negative controls were included in each plate in the rows H. Plates were incubated for 48 h at 37 °C. After this period, the protocol for measuring cell viability was performed (*see below*).

From the concentrations assayed (0X i.e., no PC, 1/2X, 1/4X and 1/8X MIC_{90}, based on previous experiments), data derived from the current experiment with 1/4X MIC_{90} concentration were used to compare with the PC MIC_{90}. Synergy was considered if $MIC_{90}(PC+SC) \leq 0.25\ (MIC_{90}(PC) + MIC_{90}(SC))$ (i.e., 4-fold reduction in efficacy).

2.22. Cell viability assays: MTT assays

One of the most often employed methods for studying cell toxicity is the 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay for cellular metabolic activity. The MTT assay is a method based on the conversion of MTT into formazan crystals only by living cells, related to active mitochondrial activity. For most cell populations, the total mitochondrial activity is related to the number of viable cells in the sample, so this assay can be used to measure the *in vitro* cytotoxic effects of drugs on cell lines and bacterial cultures (212).

Cellular cytotoxicity or viability in *B. fragilis* cultures was assessed for the two strains: enterotoxigenic and nontoxigenic. The results of the liquid MIC_{90} and synergy assays, also requiring the measurement of cell viability, were using the same protocol. The yellowish reagent MTT (Sigma Corp., St. Louis, MO) is metabolized into a purple compound only by viable cells/bacteria, while non-viable cannot metabolize it, leaving the reagent with its normal coloration. Measuring absorbance at 580 nm was possible to measure the degree of MTT conversion, which was proportional to the percentage of viable cells (i.e., viability or cytotoxicity).

As previously explained (*see liquid MIC_{90} and synergy assay sections*), the bacteria cultures were incubated with the compounds during 48 h. After this period, the cytotoxicity or cell viability was measured using the MTT assay, which was prepared by dissolving 5 mg/mL stock in water. The stock solution was protected from light and stored at 4 °C. To determine cytotoxicity, 30 µL/well of the working MTT solution (5 mg/mL MTT, 20% Tween-80) was added and incubated protected from light at 37 °C for 4 h. After that, plates were shaking cautiously and equilibrated to room temperature before reading-out. Then, they were covered with a plastic-film and the absorbance at 580 nm was measured using a Synergy HTX multi-mode plate reader (BioTek) and the Gen5 Software. Each experiment was performed in duplicate/triplicate, repeated at least two times and normalized taking untreated cells viability as a reference.

3.23. Time-kill kinetics assays

The time-kill kinetics assay (TKA) is used to study the activity of an antimicrobial agent against a bacterial strain and determine its bactericidal or bacteriostatic activity over time. This method is the most appropriate method for determining both the time- as well as concentration-dependent antimicrobial effect. Briefly, each drug is assayed in agar broth culture medium inoculated with bacterial suspensions in combination with compounds and, after time intervals, the number of living cells is calculated by the agar plate count method (111).

96-well flat bottom plates were prepared with 2X of all working conditions in each well (2X of controls and compounds alone at 10, 4, 1, 0.25, 0.1 X MIC_{90}) in a final volume of 140 µL. Then, 140 µL of the 2X inoculums for each strain (1X: $2 \cdot 10^5$ cell/well, assuming $OD_{600nm} = 0.88$ corresponds to $4 \cdot 10^8$ cell/mL) were prepared in supplemented medium and added to each well. 96-well dilution U-bottom plates (U-bottom, Tissue Culture Test Plate) were prepared by adding 180 µL of sterile saline solution (SSS) (200 µL as final volume), and agar square Petri plates (OmniTray w/Lid, non-treated sterile, polystyrene, from Thermo Scientific) for each condition were prepared. At each time point (t = 0, 2, 5, 8, 24, 48 h), 20 µL of each condition were added to the first row of the dilution plates to make 10-fold serial dilutions. 10 µL of each SSS well were transferred to the agar plates, starting from the most diluted to the more concentrated inoculums (10^{-8} to 10^{-1}), assuming that the limit of detection is 50 CFU/mL (CFU/mL: colony-forming unit/mL). Then, the agar plates were incubated under an anaerobic atmosphere for 48 h at 37 °C. After this period, the cell concentration (cell/mL) were calculated by counting the colonies at each condition. Each experiment was performed in duplicate/triplicate, repeated at least two times.

Bactericidal activity is observed when a 3log$_{10}$-fold decrease or larger in CFU (surviving bacteria) is achieved within a specified time, which is equivalent to killing ≥ 99.9% of the inoculum, while bacteriostatic activity is observed when bacteria stop reproducing, but are not necessarily dead.

3.24. *G. mellonella* waxworm larvae: purchase and maintenance

All assays related to *G. mellonella* waxworms rely on using invertebrate hosts and do not require Institutional Animal Care and Use Committee (IACUC) approval.

The waxworms themselves are cheap and easy to obtain from commercial insect suppliers, and can be housed in large numbers to allow for greater study sizes at low cost. In particular, in our case the waxworms were obtained from suppliers that do not introduce hormones, antibiotics, or other treatments to the larvae and are able to ship as well as deliver live specimens. The company from which the worms were purchased was Harkito Reptile S.L. (Madrid).

Among the most important aspects for their good condition, it is important to note that to monitor the temperature throughout the process of delivery and receipt of the worms was crucial to minimize their exposure to harsh or inappropriate environmental conditions. It should be noted that ordering during the Summer months, when the temperature usually exceeds 30 °C, decreased the viability of the larvae. Upon receipt of the larvae, it was necessary to check each container of worms to ensure that the larvae arrived viable and healthy. Larvae were considered healthy if they were completely submerged in the bedding, mobile, and had a slightly yellow/tan coloration. There should be a very small number of larvae with black spots or discolorations along the body (dead worms with advanced melanization, as well as moths, were discarded). Larvae with approximate weight of 180-250 mg were selected for *in vivo* analysis. It was necessary to work with worms of the same size, around 2 cm, which were between the 5th and 6th stage of development. The larvae were kept in the darkness in their own container, at room temperature (20-25 °C) and in a ventilated space. According to the literature, and depending on the initial state of the larvae, they can be used up to two weeks after arrival, but to guarantee optimal conditions of the larvae, all experiments (infection with *B. fragilis* and compound treatment, *see below*) were carried out on the same day of reception or on the following day. In all the experiments, the number of larvae used per condition to be tested was fixed at 10 (i.e., groups of 10 subjects), and they were kept in independent sterile plates, without bedding and without food supply, since differences in these factors could enhance variability between samples.

3.25. Preparation of *B. fragilis* cultures for *G. mellonella* infections

The manipulation of worms included proper lab attire including gloves and lab coat, and it was done in Class II Biological Safety Cabinet.

The protocol described below was based on previously published protocols for the preparation of bacterial cultures for *G. mellonella* infections, and an adaptation was made for the specific case of *B. fragilis* (213).

The bacterial cultures were refreshed in liquid media 48 h before each experiment in order to grow *B. fragilis* and have them as fresh as possible. After this growth period, 3-5 mL of liquid culture from samples were taken for both strains to prepare the cultures for infection. The cultures were centrifuged at 4,000 rpm for 5 min, supernatants were discarded and pellets were resuspended in 2-3 mL of 1X PBS. The centrifugation process was repeated at least three times, discarding the supernatant at each step, and resuspending the pellet again in the same volume of 1X PBS for washing the cells. This washing process allowed eliminating any trace from the culture medium, leaving only the pellets, consisting on bacteria dissolved in 1X PBS. After the last wash, the pellet was resuspended by pipetting, avoiding to collect the possible blood clot that remained at the bottom of the pellet, and transferred to a new tube containing 2 mL of 1X PBS. Then, 1:10/1:100 dilutions in 1X PBS were performed in order to measure the concentration by OD_{600nm} (taking into account that: OD_{600nm}= 0.88 corresponds to $4 \cdot 10^8$ CFU/mL). In this way it was possible to measure the total concentration of cells in the initial culture and dilutions in 1X PBS were made according to the working bacterial concentration needed, taking into account that each injection to larvae required 10 µL of the inoculum.

3.26. Infection of *G. mellonella* larvae with *B. fragilis* cultures

Once prepared the *B. fragilis* cultures for *G. mellonella* infections, the following steps were focused on the several steps required to perform the injections of the bacterial cultures into the waxworms.

To begin with, 1 mL of 100% ethanol and 1 mL of 1X PBS were added to two separate microcentrifuge tubes to wash the injection needle between infections. This was an important step in order to ensure that the injection needle was as clean as possible. Then, sterile 92x16 mm Petri dishes, with filter paper attached to the internal surface, were prepared in order to limit the adherence of *G. mellonella* larvae to the surface of the Petri dishes. To ensure sterile conditions, in addition to using sterile Petri dishes, they were sprayed with 70% ethanol and placed under UV light on both sides with the filter paper once adhered. According to the literature, some waxworm bedding can also be added to limit the adherence of dead larvae; however, our laboratory experience led us to the conclusion that the addition of waxworm bedding and food can cause inter-plate variability in the experiments, so our assays were carried out without the addition of waxworm bedding or food. Then, the syringe was sterilized with several washes, once loaded with a disposable needle (30G x 1/2", 0,3x13mm, BD Microlance), by aspirating and ejecting absolute ethanol, and then, several washes with 1X PBS were additionally performed. *B. fragilis* inoculation dilutions (*see B. fragilis culture preparation explained above*), were vortexed and 10 µL of culture were loaded into the syringe. It was important to see that there were no bubbles inside the syringe, as air injection may promote a higher death rate, leading to unwanted complications in data analysis. Then, the *G. mellonella* larvae container was opened and a portion of the bedding covering them was carefully removed in order to select healthy larvae. Larvae in good conditions were identified as those with good mobility, yellowish/tan coloration and absence of black pigmentation along the larval body. The size of the larvae was similar throughout the complete set of worms used in each experiment. The injection site was sterilized with 100% ethanol and then a single larva was taken and carefully held it between the index finger/heart and thumb, in a similar way to holding a pencil. The larva was turned over onto its back, so that its legs were facing up. The entire body was held with the previously mentioned fingers so that the larvae was unable to bend over or turn over. Then, the syringe was rotated so that it faced the bevel of the needle and slowly the needle was inserted into the body at the junction with the lower left proleg. The needle was fully inserted and not just pushed into the body of the larva. It was important to remark that, at this point, force was not indicated when penetrating the body, as the needle could quickly pass completely through the larva. A way to ensure that the needle had penetrated correctly was to release the larva from the hand holding it, and slowly lift the worm with the needle so that it remained hanging attached to the needle and not falling. An injection volume of 10 µL was administered and the needle was removed. Two 10 µL-injections were required in each case, i.e., 20 µL total injected volume per larva: the first one for producing (with *B. fragilis* cultures) or simulating (with 1X PBS) the infection; and the second one for treating (with compound) or simulating the compound treatment (with 1X PBS). The above steps were repeated for each larva to be infected. To prevent bacterial death/sedimentation, *B. fragilis* cultures were vortexed every few steps. As controls, 10 µL of 1X PBS were injected to a set of larvae in the case of an infection-only assay, and 10 µL of 1X PBS/DMSO in the case of a compound-treatment experiment. Each test condition was always performed with a set of 10 larvae, which were kept after inoculation/treatment in the same Petri dish until the end of the assay.

3.27. Monitoring of larvae mortality and survival analysis determination

Plates with the larvae were incubated at 37 °C in a bacterial incubator for 6 days, checking the larvae every 24 h to monitor the death/infection process induced in the worms. Mortality was initially assessed by counting those larvae that either had generalized black pigmentation throughout the body, patches or black spots which were formed at some locations on the body, or showed absence of movement. To confirm mortality, the ability of the worms to respond to stimuli was checked by using sterile clamps to carefully turn the larvae on their backs and gently touch the body. In this way, it was possible to see if they were able to turn on their own and regain their natural position, i.e., with their legs on the surface, and perform movements similar to those of uninfected worms. The absence of response observed as a lack of movement of the body or legs, as well as the inability to turn over on themselves, was also classified as infected/death. Larvae that began to develop into moths were included in the analysis as live larvae, but were removed from the Petri dishes. Survival curves for the *G. mellonella* toxicity model were plotted using the Kaplan-Meier estimator on GraphPad Prism 6.

4. RESULTS

4.1. Structural and functional characterization of BFT-3 and the role of the zinc as cofactor

4.1.1. proBFT-3 was correctly purified

One of the key aspects in the structural and functional study of proteins is to obtain the target protein with both high purity and sufficient yield to ensure enough amount of protein for subsequent studies.

During the purification process of proBFT-3 protein, an imidazole gradient was employed to separate proBFT-3 protein from the rest of the proteins that were not of our interest, after injecting the pull of proteins obtained in the bacterial lysate. The purity of proBFT-3 protein was assessed by analyzing SDS-PAGE gels from the samples at different stages of the purification process. As it can be observed (**Figure 29**), a single band with a MW of 44.4 kDa was obtained at the end of the purification steps, corresponding to the theoretical weight of proBFT-3. Furthermore, the correct identity of proBFT-3 was verified by MALDI-TOF/TOF. Therefore, it was confirmed that the protein purification was carried out correctly.

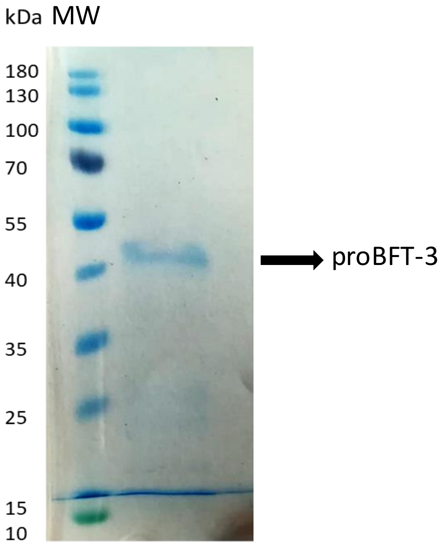


Figure 29. SDS-PAGE of pure proBFT-3. MW: molecular weight protein ladder (PageRuler^{TM} Prestained Protein Ladder).

4.1.2. proBFT-3 is a zinc-dependent conditionally disordered protein

Several biophysical techniques, such as far- and near-UV CD, DSF and both intrinsic and extrinsic probes in fluorescence spectroscopy were employed to evaluate the zinc-dependence of proBFT-3.

CD assays were performed to determine the secondary structure of proBFT-3. Consistent with the previously published proBFT-3 structure, which indicated a ~25% α-helical content (67), the far-UV CD spectrum showed a marked negative band at 220 nm (**Figure 30A**). Interestingly, the aforementioned peak increased ~11% after zinc removal facilitated by EDTA, with no significant additional modifications observed. Notably, the near-UV CD spectrum displayed non-zero values with a particular pattern (**Figure 30B**), suggesting the presence of a folded structure, and exhibited slight changes upon the zinc removal.

Furthermore, fluorescence spectra assays were performed to monitor the intrinsic fluorescence intensity of the tryptophan residues. In absence of zinc after adding EDTA, the intrinsic fluorescence signal at 350 nm was reduced by ~10% (**Figure 30C**). This decrease suggests conformational changes taking place in the local microenvironment of at least one of the four tryptophan residues within the catalytic domain of proBFT-3.

Following, the impact of zinc dissociation was evaluated using the extrinsic fluorescent probe ANS as a hydrophobic reporter (**Figure 30D**). In the presence of zinc, a fluorescence spectrum with a maximum intensity at ~505 nm was recorded, which is lower than the usual value of 535 nm for free ANS in aqueous solvent and aligns more closely with protein-bound ANS (214). Upon removal of zinc, this blue-shift further intensified to ~490 nm, and the fluorescence intensity almost doubled, indicating a potential conformational change accompanying the dissociation of the cofactor together with an increase in the exposed protein surface accessible to the solvent.

Furthermore, thermal denaturation of proBFT-3 was monitored by DSF, indicating that zinc-bound proBFT-3 unfolded in a single highly-cooperative transition, suggesting a significant folding barrier and a sharp unfolding equilibrium (**Figure 31**). The T_m of 54 °C was closely matched the previously reported value of 56 °C (67) and an unfolding enthalpy $\Delta H(T_m)$ of 101 kcal/mol was determined. Remarkably, upon zinc depletion, a substantial reduction in protein stability was observed according to the T_m and the associated unfolding enthalpy (39 °C and 74 kcal/mol, respectively). These findings indicated that the presence of zinc played an essential role in maintaining the structural stability of proBFT-3.

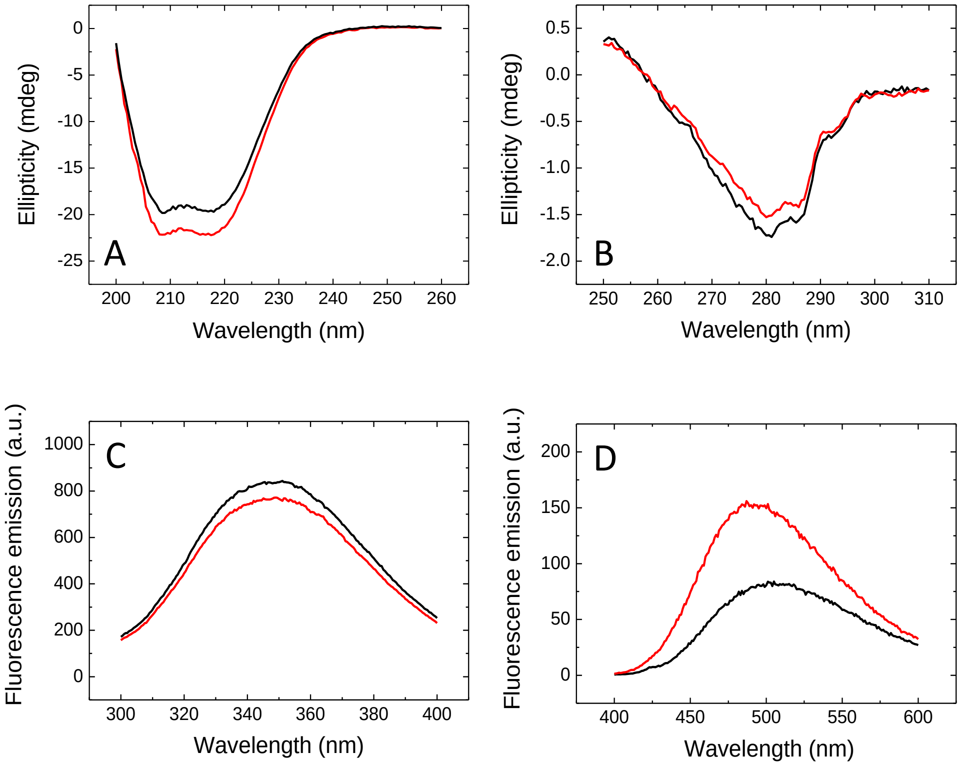


Figure 30. ProBFT-3 was subjected to spectroscopic characterization. The recorded spectra were in the presence (black) and absence (red) of zinc. The characterization included the following: (A) Far-UV CD spectra, and (B) near-UV CD spectra. Raw ellipticity data was obtained with 10 µM protein. (C) Fluorescence spectra were obtained with an excitation wavelength of 280 nm. Data was obtained with 2 µM protein. (D) ANS fluorescence spectra were recorded with an excitation wavelength of 370 nm in presence of proBFT-3. Data was obtained with 2 µM protein and 100 µM ANS.

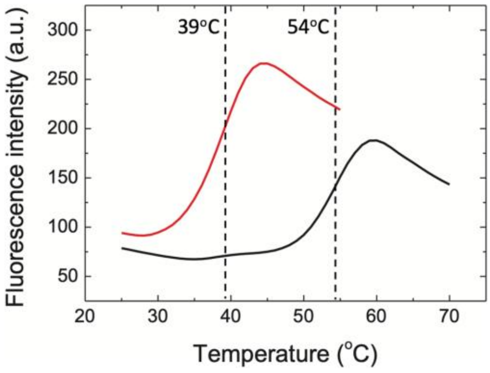


Figure 31. The thermal stability of proBFT-3 was evaluated in the presence (black) and absence (red) of zinc by DSF. Thermograms were obtained with 2 µM protein and 5X SYPRO Orange as a fluorescent probe. For zinc-free form of proBFT-3, EDTA was added. In contrast, for zinc-bound proBFT-3, no exogenous zinc was added, and the zinc concentration was that corresponding to the endogenous zinc present during the proBFT-3 purification. Consequently, a small fraction of zinc-free enzyme was observed, indicated by a small positive slope at 39 °C in the black line. The zinc bound to proBFT-3 increased the T_m from 39 °C to 54 °C, resulting in a ΔT_m of 15 °C.

4.2. Identification of novel bioactive compounds targeting the BFT-3 protein

4.2.1. Experimental screening procedure to identify potential ligands targeting the zinc-free partially disordered proBFT-3 state

After having studied the structural features of proBFT-3 structure by the biophysical characterization previously explained, the aim was to employ that information for designing and implementing an experimental screening procedure of compounds in order to, firstly, identify compounds that bind to proBFT-3, and, secondly, verify that they potentially act as inhibitors, able to block, or at least reduce, the activity of the toxin BFT-3.

In previous studies of our group, we have identified several small molecules as allosteric inhibitors of the HCV NS3 protease through a novel mechanism. In these studies, the absence of zinc produced by adding a chelator agent, such as EDTA, induced a non-native inactive partially-folded conformation of the NS3 protease that was dominated the conformational landscape. In this situation, the protein was presented as a highly unstructured state, but it still maintained an important amount of residual structure which may be stabilized by ligands. Then, ligands with the ability to bind and stabilize this inactive and partially unfolded conformation state would act as allosteric inhibitors trapping the NS3 protease into the inactive conformation (181).

Because BFT-3, like NS3, is a zinc-dependent metalloprotein, we employed a similar experimental approach to target proBFT-3. We screened an FDA-approved chemical library to identify compounds capable of stabilizing zinc-free proBFT-3 against thermal unfolding and trapping it in its zinc-free, inactive conformation. Specifically, the commercially available chemical library which was employed was the Prestwick Chemical Library (Prestwick Chemical), which contained 1,120 compounds, with known therapeutic indication, good bioavailability, and safety in humans, and covering a large chemical and pharmacological space. Thermal unfolding both in presence and absence of small molecule compounds was monitored by registering the emission signal of the extrinsic fluorescent probe SYPRO Orange by TSA.

Initial hits, from now on called MOA compounds, were identified from the screening as those compounds that substantially increased the T_m of zinc-free proBFT-3 by at least 3 °C (**Figure 32 and Table 5**), compared to the internal controls. The following figure shows the 11 compounds (MOA1-11) identified in this experimental screening from the Prestwick chemical library.

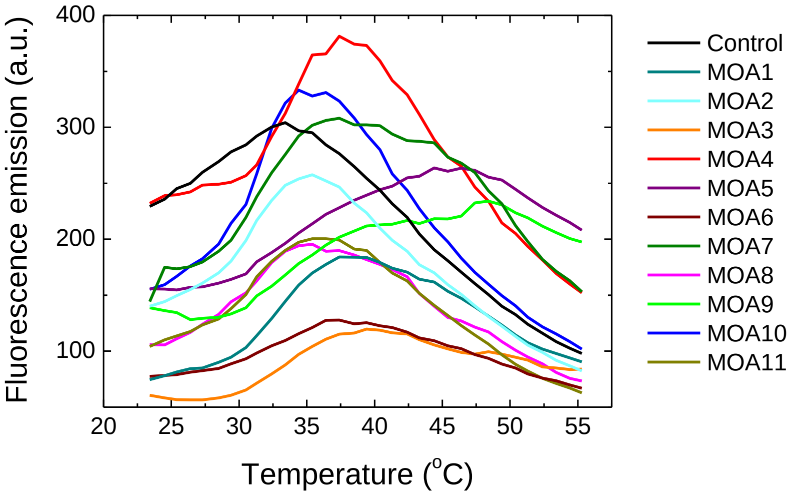


Figure 32. Experimental screening of MOA compounds by TSA. Potential ligands of zinc-free proBFT-3 were identified as those molecules increasing the T_m values of the protein by at least 3 °C. Raw fluorescence intensity data (expressed in miles) as a function of temperature is shown for compound-free protein and for protein in the presence of selected compounds.

Compound	**ΔT_m (°C)**
Control	-
MOA1	6.8
MOA2	6.0
MOA3	8.9
MOA4	6.8
MOA5	8.9
MOA6	6.8
MOA7	6.0
MOA8	6.0
MOA9	8.9
MOA10	6.0
MOA11	6.0

Table 5. ΔT_m values of proBFT-3 in presence of MOA compounds obtained by TSA.

The identification, chemical structure, MW and known therapeutic indication for each of the 11 MOA compounds is reported in the (**Table S5, Annex III**).

Chemical compounds of the chemical library were dissolved into DMSO. Therefore, the unfolding assay was carried out with a small, but necessary, residual DMSO percentage. In this assay, the presence of SYPRO Orange and DMSO could induce an additional minor destabilization, but they were in the same proportion for all samples, including the controls. Besides, control experiments were rather consistent and homogeneous (**Figure 33**).

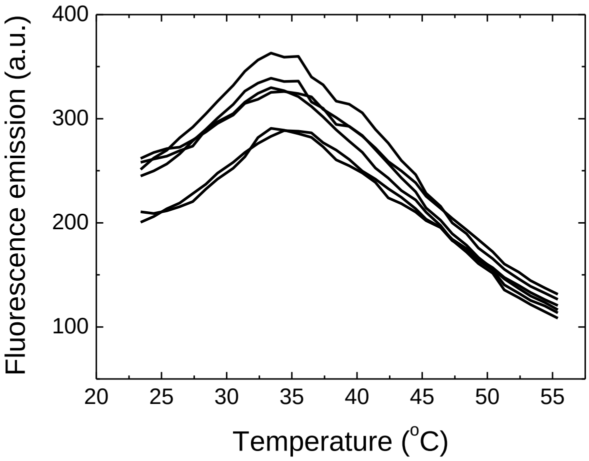


Figure 33. Control thermal unfolding curves in presence of EDTA to remove the zinc by TSA.

4.2.2. Target engagement of selected compounds by ITC

From the screening of compounds, although the initial MOA selected hits could be potential drug candidates, obviously further assays were needed to verify that these compounds, besides binding to the target protein of interest, had the desired effect on the target protein. In addition, false positives can be obtained in the screening due to compound features such as color or fluorescent properties.

After compound screening, one of the most commonly used techniques is ITC. ITC experiments were performed to characterize the binding affinity of the protein for the selected compounds, as well as the thermodynamic parameters that characterizing the interaction. Specifically, data obtained from the experimental screening, which was based on ligand-induced stabilization of proBFT-3, provided indirect evidence for protein-ligand interaction. However, in general there is no direct correlation between protein stabilization extent (i.e., ΔT_m) and ligand affinity (K_a), so the K_d were determined for the selected compounds by ITC for assessing target engagement.

Due to the extremely low solubility of MOA2, it was discarded for further testing and it was not included in the subsequent assays. Consequently, calorimetric assays were performed for all MOA compounds, except MOA2, with proBFT-3 in the presence/absence of the active site zinc. As a representation of ITC assays, calorimetric titrations of MOA4, MOA9 and MOA10 with proBFT-3 are illustrated in **Figure 34**.

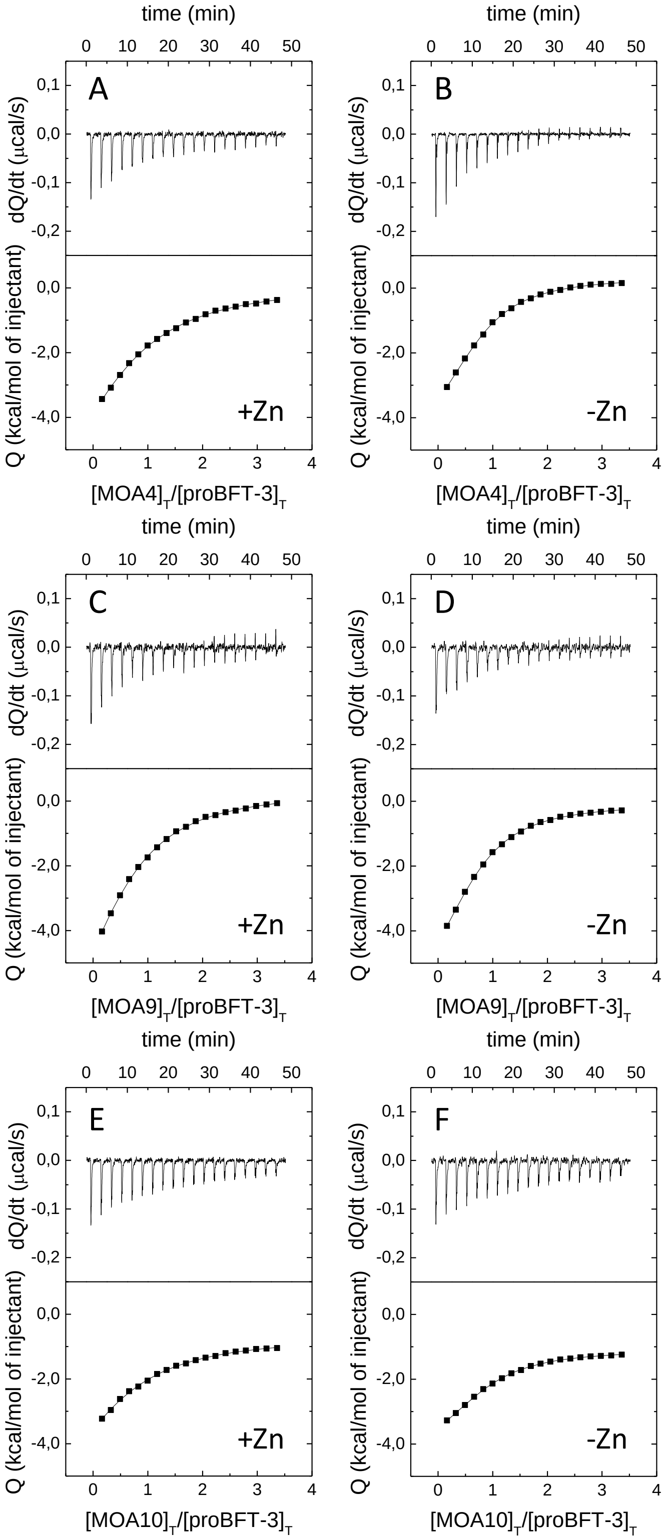


Figure 34. Interaction of selected compounds with proBFT-3. Calorimetric titrations for (A and B) MOA4, (C and D) MOA9 and (E and F) MOA10 interacting with proBFT-3, in the presence (left) and the absence (right) of zinc. Thermograms (upper panels; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherms (lower panels; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fits according to a model considering a single ligand binding site (continuous lines) are shown.

MOA4, MOA9, and MOA10 compounds exhibited K_d values in the low micromolar range, and importantly, the presence of zinc induced an approximately 2-3-fold reduction in binding affinity, which was most pronounced for MOA4 and MOA10 compounds. Interestingly, the binding enthalpies were more favorable for the zinc-bound proBFT-3, suggesting that the entropic contribution of the binding is either more favorable or less unfavorable in the zinc-free protein (**Table 6**).

Compound	**Zinc (+/-)**	**K_a (M^{-1})**	**K_d (µM)**	**ΔH (kcal/mol)**	**ΔG (kcal/mol)**	**n**
MOA4	-	$1.3 \cdot 10^5$ [$1.2 \cdot 10^5$, $1.4 \cdot 10^5$]	7.8 [7.3, 8.2]	-5.7 [-5.9, -5.5]	-7.0 [-7.5, -8.0]	0.76 [0.72, 0.74]
	+	$5.0 \cdot 10^4$ [$4.4 \cdot 10^4$, $5.6 \cdot 10^4$]	20 [18, 23]	-9.1 [-10.5, -8.1]	-6.4 [-6.9, -5.9]	0.73 [0.71, 0.80]
MOA9	-	$1.2 \cdot 10^5$ [$1.1 \cdot 10^5$, $1.2 \cdot 10^5$]	8.5 [8.0, 8.9]	-6.8 [-7.1, -6.6]	-6.9 [-7.4, -6.4]	0.67 [0.66, 0.68]
	+	$6.9 \cdot 10^4$ [$6.3 \cdot 10^4$, $7.6 \cdot 10^4$]	14 [13, 16]	-10.2 [-11.2, -9.4]	-6.6 [-7.1, -6.1]	0.66 [0.62, 0.69]
MOA10	-	$1.2 \cdot 10^5$ [$1.1 \cdot 10^5$, $1.2 \cdot 10^5$]	8.7 [8.2, 9.2]	-3.6 [-3.7, -3.5]	-6.9 [-7.4, -6.4]	0.83 [0.82, 0.85]
	+	$4.6 \cdot 10^4$ [$3.7 \cdot 10^4$, $5.6 \cdot 10^4$]	22 [18, 27]	-7.5 [-8.8, -6.1]	-6.4 [-6.9, -5.9]	0.67 [0.55, 0.75]

Table 6. Thermodynamic parameters of the proBFT-3-compound complexes obtained by ITC at 25 °C and pH 7.4. The uncertainty in the estimation of the binding parameters is reported as the confidence interval at a statistical significance of 68%, shown in square brackets below each parameter. Association constant, K_a; dissociation constant, K_d; binding enthalpy, ΔH; binding stoichiometry (or percentage of binding-competent protein).

Additional experiments done with trypsin-activated BFT-3 in the presence/absence of the active site zinc provided similar binding affinities and similar influence of the zinc cofactor (**Table 7**). Noticeably, the binding enthalpies were more favorable (i.e., more negative) than those for proBFT-3, and, therefore, the binding entropies were less favorable or more unfavorable than those determined for proBFT-3.

Compound	Zinc (+/-)	K_a (M^{-1})	K_d (μM)	ΔH (kcal/mol)	ΔG (kcal/mol)	n
MOA4	-	$1.2 \cdot 10^5$ [$1.1 \cdot 10^5$, $1.5 \cdot 10^5$]	8.1 [6.7, 9.1]	-10.7 [-11.3, -10.2]	-6.9 [-7.4, -6.4]	0.69 [0.64, 0.73]
	+	$6.1 \cdot 10^4$ [$5.0 \cdot 10^4$, $7.3 \cdot 10^4$]	16 [14, 20]	-13.1 [-14.4, -11.0]	-6.5 [-7.0, -6.0]	0.64 [0.57, 0.70]
MOA9	-	$1.1 \cdot 10^5$ [$9.8 \cdot 10^4$, $1.3 \cdot 10^5$]	9.1 [7.7, 10]	-9.6 [-10.1, -9.1]	-6.9 [-7.4, -6.4]	0.64 [0.62, 0.66]
	+	$6.8 \cdot 10^4$ [$6.0 \cdot 10^4$, $8.5 \cdot 10^4$]	15 [12, 17]	-16.2 [-17.4, -15.1]	-6.6 [-7.1, -6.1]	0.65 [0.60, 0.72]
MOA10	-	$1.0 \cdot 10^5$ [$9.4 \cdot 10^4$, $1.3 \cdot 10^5$]	10 [7.7, 11]	-7.5 [-7.8, -7.1]	-6.8 [-7.3, -6.3]	0.73 [0.71, 0.75]
	+	$5.2 \cdot 10^4$ [$4.1 \cdot 10^4$, $6.1 \cdot 10^4$]	19 [16, 24]	-11.6 [-13.0, -10.3]	-6.4 [-6.9, -5.9]	0.71 [0.68, 0.76]

Table 7. Thermodynamic parameters of the activated BFT-3-compound complexes obtained by ITC at 25 °C and pH 7.4. The uncertainty in the estimation of the binding parameters is reported as the confidence interval at a statistical significance of 68%, shown in square brackets below each parameter. Association constant, K_a; dissociation constant, K_d; binding enthalpy, ΔH; binding stoichiometry (or percentage of binding-competent protein).

In the case of MOA1, MOA6 and MOA8 compounds, calorimetric titrations were inconclusive. However, this aspect did not mean that these compounds did not bind to the protein, but that under the specific experimental conditions employed in the assays (i.e., protein/compound concentration, temperature, buffer...), the interaction was not detected. Therefore, calorimetric titrations could have been obtained for estimating their affinity for proBFT-3 under other experimental conditions.

Hence, these compounds were not excluded for the subsequent assays because additional complementary techniques to ITC were required before eliminating any compound from consideration (i.e., a single technique is not sufficient as a selection criterion).

4.3. Effect of the selected compounds on the BFT-3 protein using *in vitro* and *in vivo* systems

4.3.1. The selected compounds inhibited the enzymatic activity of BFT-3 *in vitro*

After confirming the binding of most of the MOA compounds to the proBFT-3 protein, we assessed whether this binding led to any modifications in the activity protease of the protein. Notably, as previously mentioned, MOA1, MOA6 and MOA8 did not bind to proBFT-3 in ITC assays. Nevertheless, we also evaluated these compounds as potential inhibitors of the BFT-3 enzymatic activity *in vitro*.

To assess the potential inhibitory effect of the MOA compounds on the BFT-3 proteolytic activity, we utilized the commercial EnzChek Protease Assay kit. As previously detailed in the Materials and Methods section, the proteolytic activity of BFT-3 results in the cleavage of the quenched BODIPY FL casein substrate, causing an increase in fluorescence intensity that is directly proportional to the enzymatic activity. Consequently, higher protease activity in the sample corresponds to an increased fluorescence.

As proBFT-3 does not exhibit activity in the proprotein form, the protein was activated with trypsin leading to the BFT-3 mature form. The activity determined for trypsin-activated BFT-3, proBFT-3 alone and trypsin alone (used as a positive control) was compared to evaluate their potential contributions to the overall signal (for example, to take into account possible auto-activation of proBFT-3).

As expected, an increase in fluorescence over time was detected for BFT-3, but not for proBFT-3, and in presence of all MOA compounds, a dose-dependent inhibition was observed (**Figure 35**).

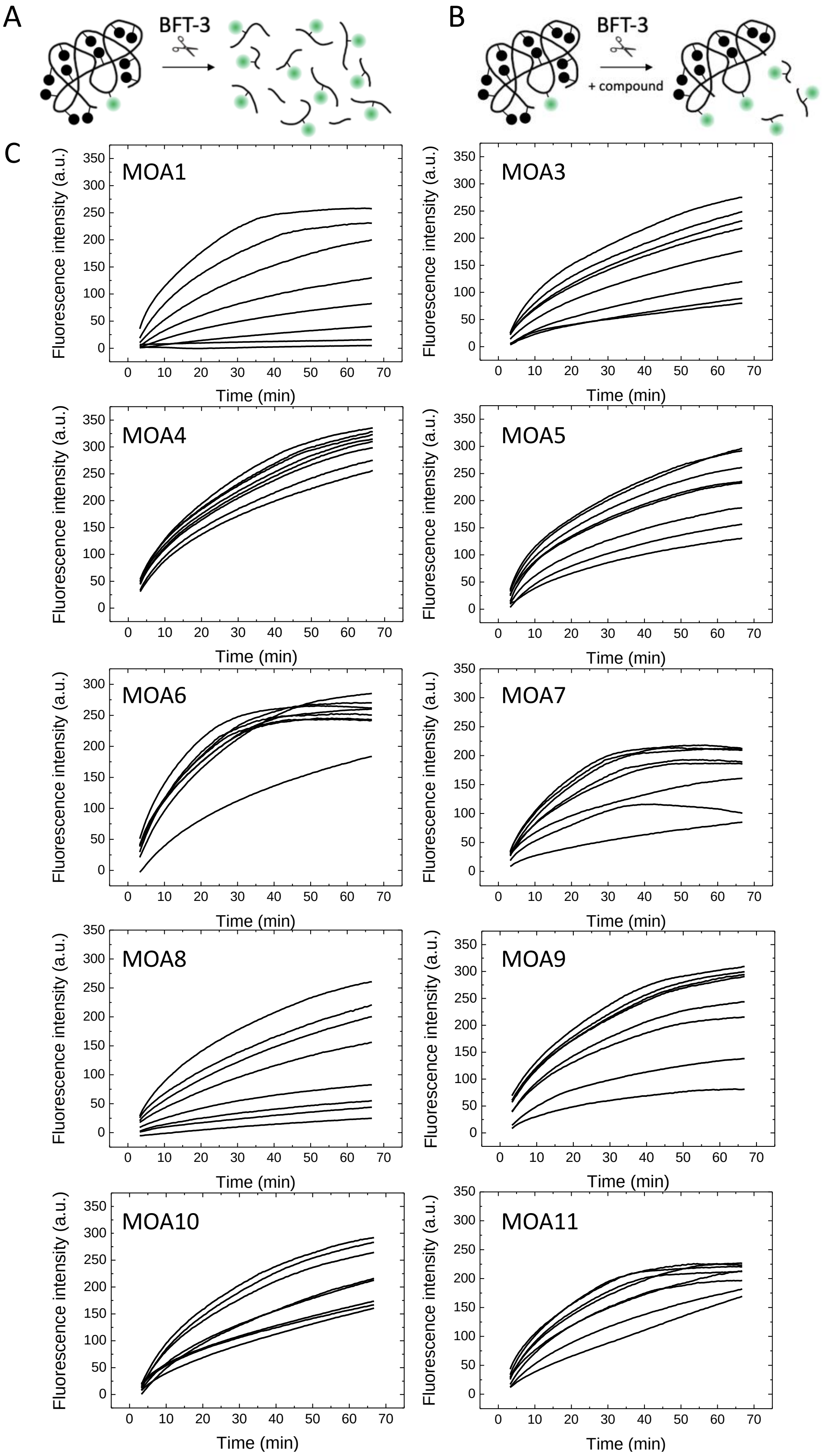


Figure 35. Proteolytic activity of BFT-3 was measured *in vitro* and the inhibitory effect of selected MOA compounds was evaluated. (A-B) BODIPY FL casein was employed as a fluorescent substrate for trypsin-activated BFT-3. Due to proteolytic processing by BFT-3, a loss of reciprocal

fluorophore quenching occurred, resulting in an increase of fluorescence signal over time. (A) Maximal activity was observed with BFT-3; (B) The addition of an inhibitor compound reduced the hydrolytic rate, observing a lower fluorescence signal intensity. (C) Increasing the concentration of the MOA compounds (2-fold serial dilutions from 1 to 0 mM) decreased the activity of BFT-3 in a dose-dependent manner.

An important aspect was that, proBFT-3 showed no activity, while trypsin showed some activity at the concentration employed, although considerably lower than that of BFT-3. To account for this, trypsin's contribution was subtracted from the total signal of trypsin-activated BFT-3. This subtraction was performed for each trypsin-activated BFT-3 in combination with MOA compounds. **Figure 36** illustrates the proteolytic activity of proteases with or without MOA10 as an example, with similar observations made for all MOA compounds (data not shown).

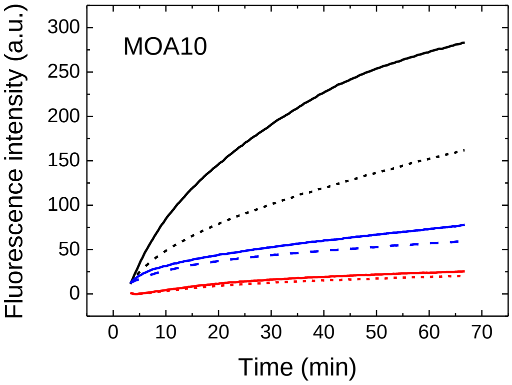


Figure 36. The proteolytic activity of proteases was measured. Loss of reciprocal fluorophore quenching due to proteolytic processing resulted in an increase of fluorescence over time. The time evolution for the substrate-associated fluorescence signal was measured *in vitro* after adding: inactive zymogen proBFT-3 (red), trypsin-activated BFT-3 (black), and trypsin (blue). The discontinuous lines show the effect of compound MOA10 at 1 mM concentration. proBFT-3 showed a residual activity likely due to self-activation. Trypsin showed some activity, but much lower than that of BFT-3. The inhibition effect of MOA10 was considerable for BFT-3, but much smaller against trypsin.

4.3.2. MOA compounds showed little or no cytotoxic effect on HT-29 and HeLa cells

As mentioned above, the drug discovery process requires a great effort in which different areas of research and disciplines play important roles. After performing the initial *in vitro* steps to identify compounds active against the selected target proteins, it is necessary to evaluate both their pharmacokinetics and toxicity in biological systems. It is important to note that a compound that is highly active and effective *in vitro* is not necessarily the ideal candidate *in vivo*, since the pharmacological action of a compound depends not only on its interaction with the biological target, but also on its bioavailability, which is closely related to its solubility and its ability to cross biological membranes or to its ability to interact with other proteins or to its susceptibility to metabolic degradation. It was therefore necessary to evaluate the cytotoxicity of the compounds identified in the screening, to ensure a rapid and early elimination of those compounds with undesirable side effects in future clinical trials, and to ensure that only those compounds with high and real potential as drug candidates were selected.

Cytotoxicity of the selected compounds of the experimental screening was tested on cell cultures. Concretely, MOA compounds were tested on both HT-29 and HeLa cells. Cell survival or viability was evaluated increasing compound concentrations, from 0 µM to 400 µM (at the same DMSO final concentration) to determine the cytotoxic concentration 50 (CC_{50}) (**Figure 37**).

CC_{50} is defined as the compound concentration at which 50% of the cells are able to survive (survival rate is quantified as the metabolic activity rate compared to control values, for which only DMSO were added).

Importantly, most of the MOA compounds showed little or no cytotoxic effect, with high CC_{50} values and suggesting a potential appropriate therapeutic window for inhibition (**Table 8**). MOA5, MOA7 and MOA8 were the more cytotoxic compounds in both cell lines used, due to the cell viability decreasing at the concentrations employed. Furthermore, in most cases it was not possible to observe differential toxic effects depending on the cell line.

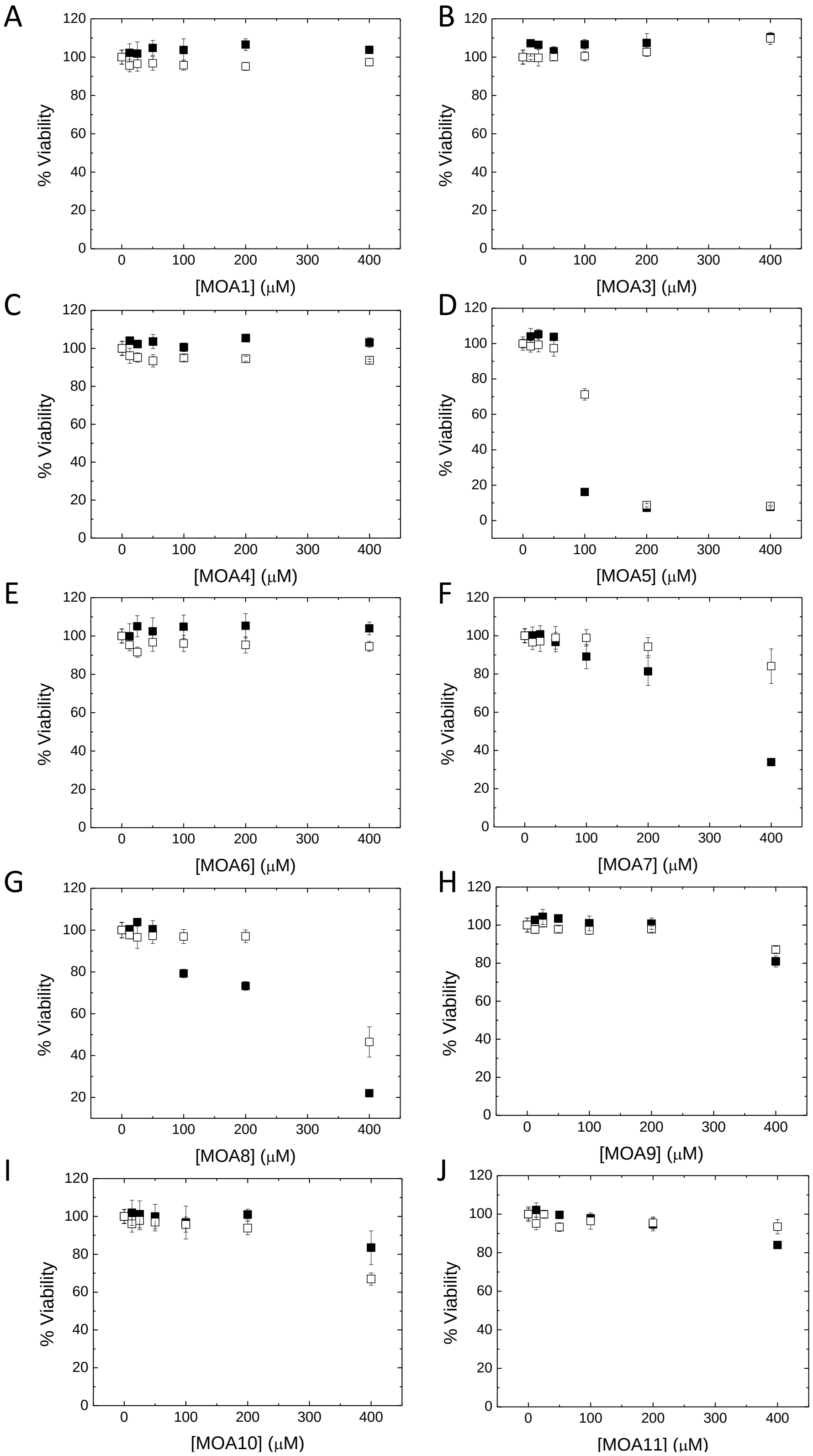


Figure 37. The cytotoxicity of MOA compounds against HT-29 and HeLa cell lines was evaluated. Cell viability values for HT-29 (open squares) and HeLa cells (filled squares) determined by CellTiter 96® assay after 48 h of incubation with increasing concentrations of (A) MOA1, (B): MOA3, (C) MOA4, (D) MOA5, (E) MOA6, (F) MOA7, (G) MOA8, (H) MOA9, (I) MOA10, and (J) MOA11. All data are presented as the average ± SD of three biological replicates, performing the assay twice with technical replicates.

Compound	CC_{50} HeLa cells (μM)	CC_{50} HT-29 cells (μM)
MOA1	> 400	> 400
MOA3	> 400	> 400
MOA4	> 400	> 400
MOA5	80	150
MOA6	> 400	> 400
MOA7	350	> 400
MOA8	300	380
MOA9	> 400	> 400
MOA10	> 400	> 400
MOA11	> 400	> 400

Table 8. CC_{50} for the selected MOA compounds in two different cell types: HeLa and HT-29 cells.

4.3.3. Optical microscopy assays were not determinant in evaluating the effect of MOA compounds

As previously indicated in the introduction of Chapter I, BFT-3 produces morphological changes, including cell rounding and dissolution of cell clusters after being added to cell cultures. For this reason, the effect produced after the addition of different concentrations of proBFT-3 in HT-29 cells was observed by microscopy. When cells were treated at the higher concentration of proBFT-3, the morphological aspect was worse than that for a lower concentration of the toxin, indicating that there was a dose-dependent effect (**Figure 38**). These assays also verified that, although purified in the proprotein form (proBFT-3), proteases in the cell culture medium caused the transformation to the mature form of the protein (BFT-3), and thus activated, with cytotoxic effects on the cells.

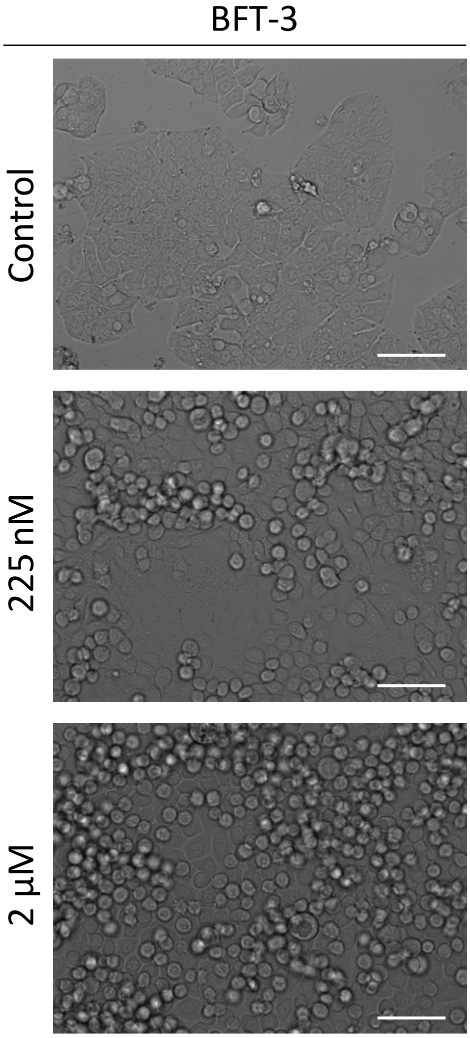


Figure 38. HT-29 cells were treated with 0 nM (control), 225 nM or 2 µM BFT-3. After adding proBFT-3, morphological changes occurred in cells comparing to control, including cell rounding and dissolution of cell clusters. Scale bar: 100 µM.

It was also evaluated whether, in addition to proBFT-3 (at 775 nM), adding the MOA compounds at different concentrations to HT-29 cells, a lower level of cell damage could be observed, thus indicating that the compounds were inhibiting the protein *in vitro* in cell cultures (**Figure 39**).

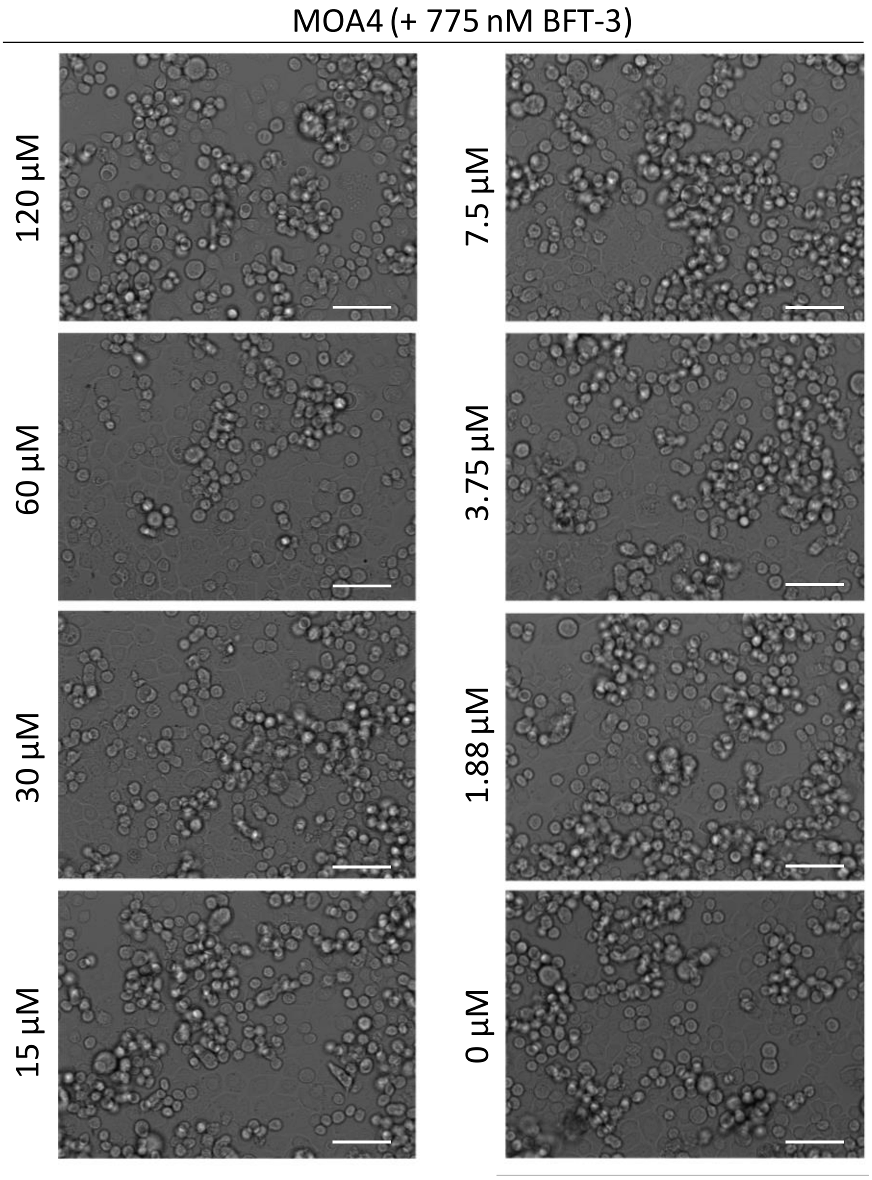


Figure 39. Representative image of HT-29 cells treated with MOA4 compound. In the image, example of two-folded serial dilutions from 120 to 0 µM MOA4 and 775 nM BFT-3. Images for other MOA treatment are not shown, since the results obtained were similar. Scale bar: 100 µM.

However, although under some condition some degree of inhibition of the proteolytic activity by the compounds was apparent, no clear results could be obtained with this technique. Moreover, since the morphological changes were observed visually, i.e., it was a subjective and qualitative evaluation, analyzing the percentage of inhibition of the compounds by this technique was not very accurate. Therefore, microscopy assays were not sufficient to determine the potential inhibitory effect of MOA compounds *in vitro* in cell cultures, and other complementary techniques were needed.

4.3.4. Western blot assays revealed that selected compounds inhibit E-cadherin processing in cell cultures *in vitro*

It was previously explained in the Introduction the proposed mechanism by which BFT-3 produces the cleavage of E-cadherin (~120 kDa) after its binding to a host CEC receptor. First, the 80 kDa extracellular E-cadherin ectodomain is hydrolyzed, and the 33 kDa and 28 kDa fragments are hydrolyzed later (75,76,81).

Western blot assays were performed to evaluate whether the selected MOA compounds were able to hamper E-cadherin hydrolysis catalyzed by BFT-3 in HT-29 cells. After the treatment of HT-29 cells with proBFT-3 and the selected compounds, the proBFT-3 should have been activated to the BFT-3 mature protein by proteases, such as trypsin, which were present in the cell culture. Intact E-cadherin, meaning that it contained both ecto- and endo-domains, would be observed as a single band at ~120 kDa. If a scission was produced, processed E-cadherin would be observed as a band at ~80 kDa, corresponding to the E-cadherin ectodomain.

Due to the fact that in the cytotoxic assays most of the selected MOA compounds showed little toxicity up to 400 µM (MOA5, MOA7 and MOA8 were the more cytotoxic, being MOA5 the most cytotoxic compound), and they were safe in cells at concentrations up to 200 µM (with the exception of MOA5, being therefore this compound eliminated for further assays), cells were treated with them to analyze if they had the capacity to inhibit the processing of E-cadherin without producing toxic side effects. After the exposition to the toxin, cells were treated with compounds MOA1-MOA11 (excluding MOA2 and MOA5) at 120 µM, and in some cases, such as employing MOA4, MOA9, MOA10 and MOA11, the presence of those compounds reduced the levels of the E-cadherin processing (**Figure 40**), maintaining most of E-cadherin in its unprocessed form (MW 120 kDa). The addition of other MOA compounds, such as MOA6 and MOA7, inhibited E-cadherin processing only in some assays, observing less consistent results that might depend on the good quality status of the cells. However, the presence of MOA1, MOA3 and MOA8 did not maintain the E-cadherin in its unprocessed form, meaning that these compounds may not have an inhibition effect of BFT-3 protease activity in HT-29 cell cultures. Therefore, MOA1, MOA3 and MOA8 were eliminated for further assays. Of note, GAPDH (~36 kDa) was used as a housekeeping gene for comparison purposes as well as for evaluating a suitable quality control of the western blot assays.

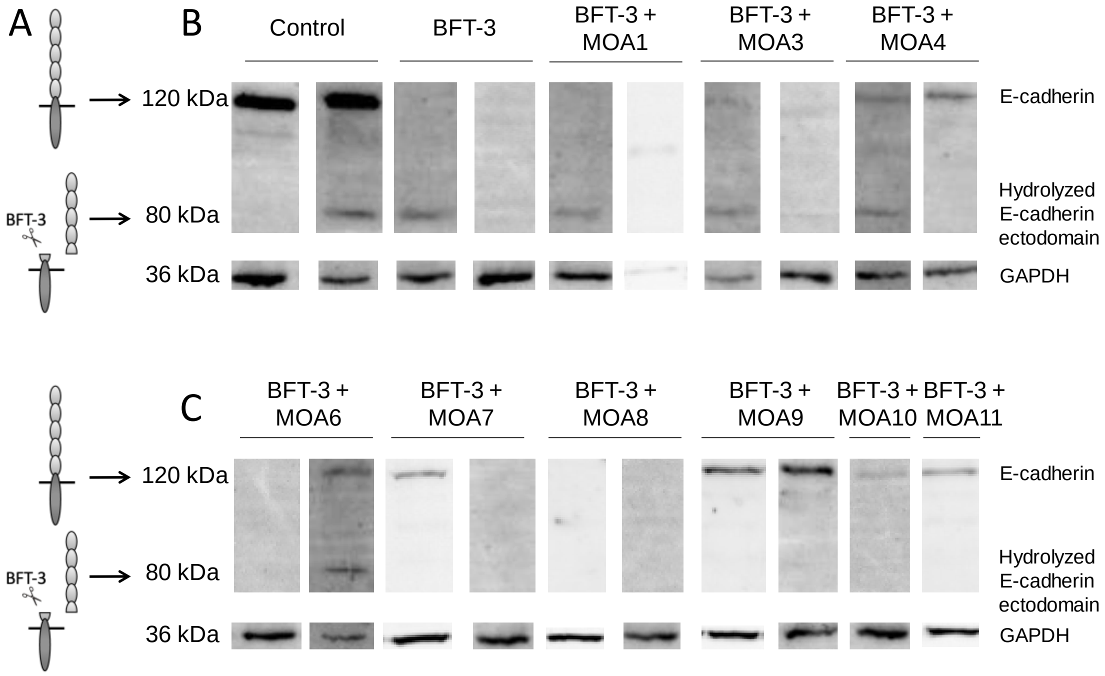

A
B
Control
BFT-3
BFT-3 + MOA1
BFT-3 + MOA3
BFT-3 + MOA4
120 kDa
80 kDa
36 kDa
BFT-3
E-cadherin
Hydrolyzed E-cadherin ectodomain
GAPDH
C
BFT-3 + MOA6
BFT-3 + MOA7
BFT-3 + MOA8
BFT-3 + MOA9
BFT-3 + MOA10
BFT-3 + MOA11
120 kDa
80 kDa
36 kDa
BFT-3
E-cadherin
Hydrolyzed E-cadherin ectodomain
GAPDH

Figure 40. The BFT-3 proteolytic activity and inhibitory effect of MOA compounds were monitored in a HT-29 cell-based assay. (A) BFT-3 processing of E-cadherin (MW 120 kDa) resulted in the release of its ectodomain (MW 80 kDa). (B-C) In the absence of BFT-3, no degradation of E-cadherin was observed (control), whereas the presence of BFT-3 led to complete degradation of full-length E-cadherin, while the ectodomain was still partly detectable. The presence of some MOA compounds (MOA4, MOA9, MOA10 and MOA11) at 120 µM inhibited E-cadherin processing, as evident by detected western blot bands at 120 kDa. Others, such as MOA6 and MOA7, inhibited E-cadherin processing just in some assays; while MOA1, MOA3 and MOA8 did not inhibited this E-cadherin processing. When proBFT-3 was added to the assay, proteases that were present in the cell culture medium, as well as self-activation, resulted in activated BFT-3 protein under these conditions.

4.3.5. Structure determination of inhibitory complexes by crystallographic assays

Once it was established that the selected MOA compounds interacted with BFT-3 and had an effect on the protein by several techniques, it was important to perform protein crystallographic assays to better understand such interaction and find out which residues were involved in the ligand-protein binding.

Taking into account the previous results from the western blot assays, co-crystallization of proBFT-3 in the presence of MOA4, MOA6, MOA7, MOA9, MOA10 and MOA11 compounds was planned, due to the fact that, as previously mentioned, MOA1, MOA3 and MOA8 were eliminated for further assays. The co-crystallization of these most promising compounds was successful for three compounds: MOA4, MOA9 and MOA10.

Structures files of BFT-3 protein in PDB format were obtained from crystallographic assays. The native protein structure was firstly deposited in the PDB (code: 3P24) (67) with a resolution of 1.80 Å some years ago. Specifically, in the frame of this Doctoral Thesis, five crystal structures of proBFT-3, unbound and in complex with three of these identified inhibitors were successfully determined, and they were deposited in the PDB database with the following PDB entries: 7PND for unbound proBFT-3, 7POL for proBFT-3 in complex with MOA4, 7POO for proBFT-3 in complex with MOA9 (ORTH space group), 7POQ for proBFT-3 in complex with MOA9 (TETR space group), and 7POU for proBFT-3 in complex with MOA10. These new structures were obtained in the ALBA Synchrotron (Barcelona) for MOA4 and MOA9-ORTH, and in the beamline Diamond i04-1 (Oxfordshire) NATI, MOA9-TETR and MOA10.

In addition, these findings revealed the presence of an, until now unknown, distal exosite in BFT-3, which is defined as a secondary binding site remote from the active site on a given protein and that is responsible for inhibitor (or general effector) binding. This exosite in BFT-3 is located on the surface of the CatD opposed to the active site cleft. A more general exosite representation was obtained by overlaying the MOA compounds bound to the protein in these crystallographic assays, after superposing the respective proBFT-3 moieties (**Figure 41**). The corresponding molecular representations of all five determined X-ray structures, each containing two protomers per asymmetric unit, are shown in **Figure 42**.

Before presenting more in detail the results obtained with MOA compounds in complex with proBFT-3 in the crystallographic assays, it was necessary to deeply analyze some important aspects of the proBFT-3, related to its amino acid residues and its conformation.

As previously mentioned in the introduction, proBFT-3 is a two-lobed zinc-dependent MP zymogen which encompasses various domains. proBFT-3 has a signal peptide, a PD, and a comparably lengthy CatD. Notably, the two globular entities of the PD and CatD are connected by an inhibitory C-terminal linker segment located within the PD, in which the D^{194} of the linker binds to the catalytic zinc ion, thereby displacing a water molecule present in the active site and rendering the protease inactive in its zymogenic form. The catalytic site of proBFT-3 is formed by a consensus zinc-binding motif, denoted as $HE^{349}XXHXXGXXH$, containing three zinc-coordinating His residues and a Glu residue. Additionally, the active site is further complemented by a loop segment and a tight 1,4-β-turn centered on M^{366} (**Figure 41**).

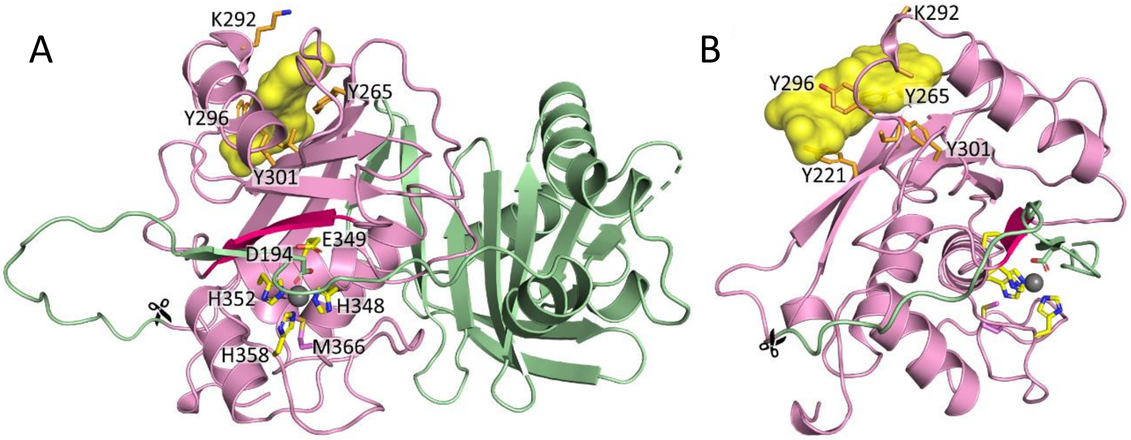


Figure 41. The crystallographic structure of proBFT-3 in complex with the identified inhibitors is represented. (A) Overall structure in front view and (B) side view of proBFT-3. The unbound proBFT-3 structure (PDB 7PND) was represented by green (PD) and pink ribbons (CatD), with α-helices and β-strands as coiled ribbons and flat arrows, respectively. The active site residues were illustrated as sticks with yellow carbons, the catalytic zinc was represented as a grey sphere, and the M^{366} of the hydrophobic basement was indicated by violet carbon atoms. Notably, the linker segment of the PD bound to the CatD in reverse direction compared to a substrate, and the aspartate-switch D^{194} was depicted as stick mode with green carbons. The exosite, depicted as a yellow semi-transparent surface was located at the back rear of the CatD, approximately 25 Å away from the active site. The exosite was obtained by overlaying the MOA compounds that bound to the protein (PDB entries: 7POL, 7POO, 7POQ and 7POU), after superposing the respective proBFT-3 moieties. Key inhibitor-binding residues were shown as sticks with orange carbons. The upper-rim strand of the CatD, which guided the substrate, was represented in hot pink, and the activation site of native BFT-3 was indicated with a scissor symbol. The polypeptide chain was interrupted between the segment S^{161}-G^{167} of the PD, as denoted by a dashed line. Of note, in the side view, residues V^{34}-N^{185} of the PD were omitted for a better clarity.

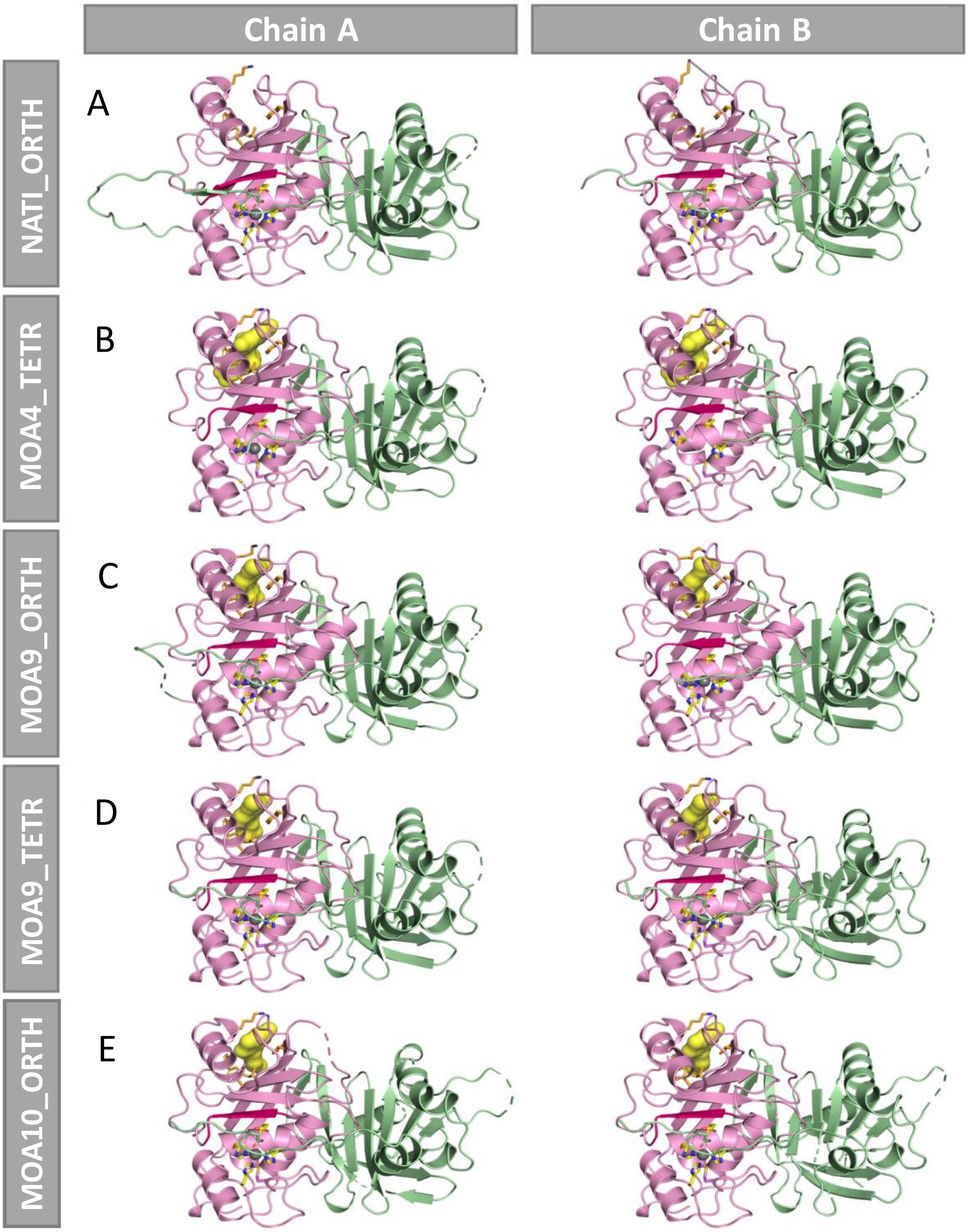

Chain A
Chain B
NATI_ORTH
A
MOA4_TETR
B
MOA9_ORTH
C
MOA9_TETR
D
MOA10_ORTH
E

Figure 42. An overview representation of all proBFT-3 structures that were determined in crystallographic assays is shown. The left and right panels showed the two molecules of the asymmetric unit. The same orientation and settings as Figure 41 were employed, i.e., with the active site and key ligand binding residues in stick mode, and the bound exosite inhibitors represented by a yellow surface. A dashed grey line indicates the distance between $K^{292}N\zeta$ and $T^{266}C\alpha$ for chain B in both unligated and MOA4-complexed proBFT-3 structures.

It is noteworthy that all compounds were bound in a similar orientation within a prominent pocket located at the top rear of the CatD. In this pocket, the inhibitors inserted in a wedge-like region between the end of the so-called “adamalysin helix” (α5) and the first three strands (β12, β13, β14) of the β-sheet of the upper subdomain (**Figure 43 and Figure 42, left and a middle panels**). In the two protomer complexes present in all crystallographic asymmetric units, this binding pattern was consistently observed.

An important finding was that, while for MOA9 (307.35 Da) and MOA10 (302.29 Da) one inhibitor molecule was found at the exosite with interaction areas of 382 and 364 Å^2, respectively, two molecules of the smaller MOA4 (261.25 Da) were found, resulting in an increased interface from 304 Å^2 for a single MOA4 molecule to 497 Å^2. Interestingly, the first MOA4 molecule was observed to be deeply buried within the protein cavity, with only 14% of its surface being accessible to solvent. On the other hand, the second MOA4 molecule was found to be situated further outside, with 43% of its surface being solvent accessible. In comparison, for MOA9 and MOA10, 76% of the surfaces were buried by proBFT-3, which was highly similar to the combined 71% burial observed for the two MOA4 molecules. However, despite the observed differences in burial and solvent accessibility, the ITC data did not provide evidence for the second MOA4 molecule, most likely attributed to the lower inhibitor concentrations used in the calorimetric experiment, suggesting a much lower affinity for the second binding event.

Another important aspect was that the key ligand interacting residues in proBFT-3 involved four Tyr residues and a Lys residue (K^{292}). Specifically, Tyr Y^{221}, Y^{265} and Y^{301} were located on β-strands β12, β13 and β14, respectively, while Y^{296} was positioned in the loop after helix α5. Together, they collectively defined the ligand binding pocket and literally created a sandwich-like arrangement around the bound ligands, with the Lys closing the pocket in from the top (**Figure 43, left panel**). It is important to note that, while the Tyr residues exhibited almost perfect superimposition between both the non-ligated and complexed structures, the Lys sidechain closed the distance to the opposing side of the exosite cleft by approximately 2 Å (measured between K^{292}Nζ and T^{266}Cα; see e.g., **Figure 42**). Moreover, while the Tyr residues contributed with both hydrophobic and parallel π-stacking interactions to inhibitor binding, K^{292} played a significant role by providing, in addition to hydrophobic interactions, a salt bridge and a hydrogen bond in the MOA4 and MOA10 complexes, respectively (**Figure 43**).

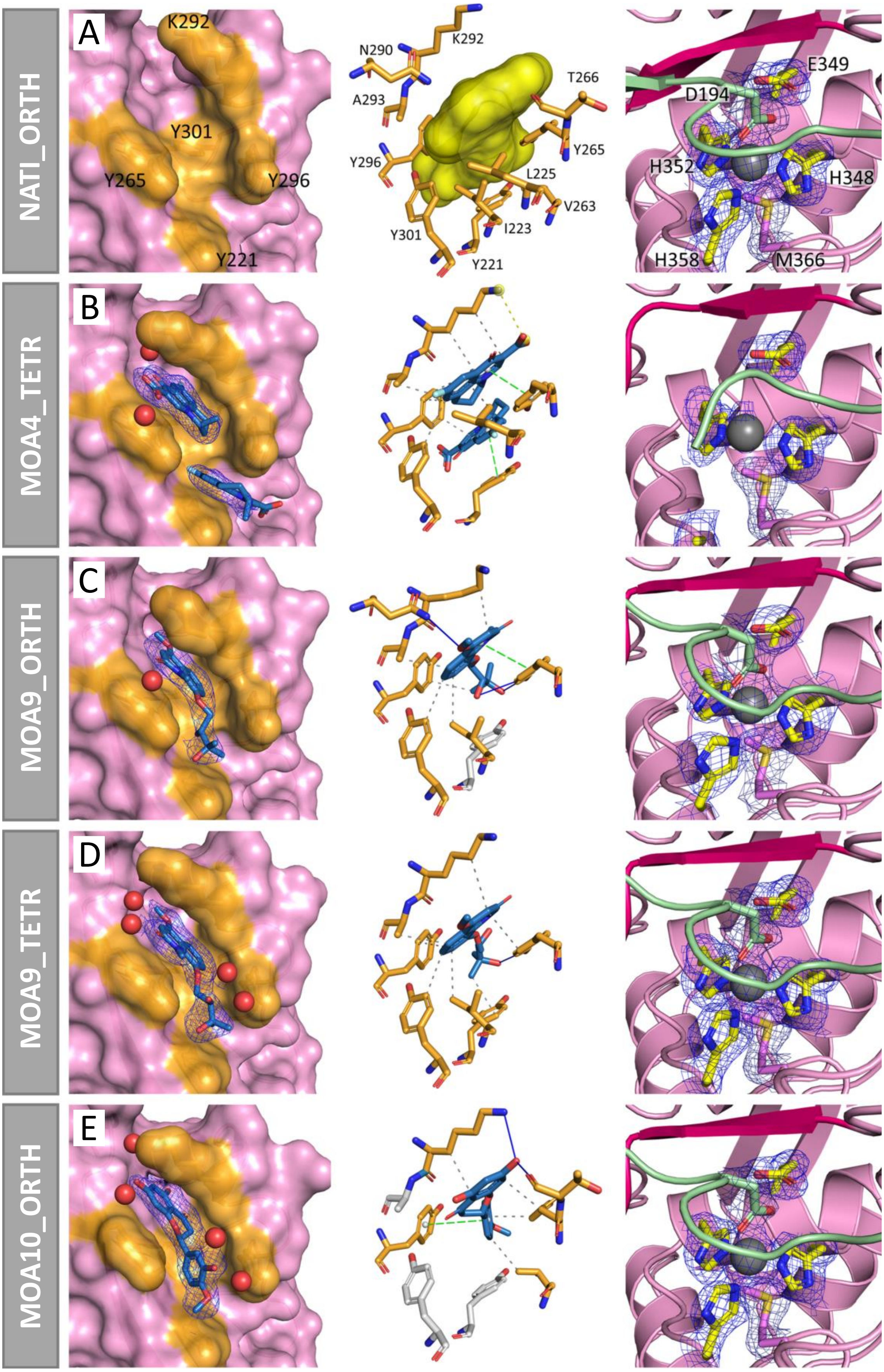

NATI_ORTH
A
K292
Y301
Y265
Y296
Y221
N290
K292
A293
T266
Y265
Y296
L225
V263
I223
Y301
Y221
E349
D194
H352
H348
H358
M366
MOA4_TETR
B
MOA9_ORTH
C
MOA9_TETR
D
MOA10_ORTH
E

Figure 43. Binding of inhibitors to the newly discovered exosite on proBFT-3 are represented. The left image displays the exosite in the distinct structures where the inhibitors bound, in the middle is represented the protein-ligand interaction network and on the right the active site of the CatD is shown. Of note, each X-ray structure represented chain A of the crystallographic asymmetric unit (the corresponding second protein molecule, i.e., chain B, is shown in Figure 44, with the same settings and orientations as presented here). In the left image, a back view of the ligand-binding exosite is shown. The proBFT-3 protein was depicted as a pink semi-transparent surface, and the key ligand-binding residues were represented as sticks with orange carbons. The respective inhibitor was displayed in sky-blue and stick mode, with the corresponding electron density map (2Fo-Fc) contoured at 1.0 sigma, and utilizing a carve radius of 2.0 Å. Water molecules that were within hydrogen bond-forming distance (≤3.4 Å) were illustrated as red spheres. In the middle section, the protein-ligand interaction networks at the inhibitory exosite are presented. The orientation was similar to the overview structure depicted in Figure 41, and rotated approximately 180° away compared to the view of the left figures. The inhibitors were represented in atom-color mode, with carbons shown in sky blue. In (A), the exosite of unligated proBFT-3 was displayed, but complemented with the cumulative surface of all crystallized inhibitors, shown in yellow, as in Figure 41. (B) For MOA4, two copies of the inhibitor were found at each exosite. Regarding MOA9, two crystal forms were identified: (C) ORTH space group and (D) TETR space group. (E) The ligand interaction network for MOA10 was represented. All protein-ligand interactions were calculated using the PLIP webserver 43, and were displayed as dashed lines in gray, green, and yellow, representing hydrophobic interactions, parallel π-stacking, and salt bridges, respectively, while hydrogen bonds were represented as solid blue lines. In the right part of the image, active site views of proBFT-3 in both non-ligated and inhibitor-complexed structures were presented. The orientation was similar to that shown in Figure 41, with the same key active site residues displayed in stick mode. The molecular representation was overlaid with the corresponding electron density map (2Fo-Fc), which was contoured at 1.0 sigma, and using a carve radius of 2.0 Å. It is important to note that the active site of MOA4-complexed proBFT-3 was completely perturbed, highlighted by a low zinc occupancy and the absence of electron density for the third proteinaceous zinc ligand (H^{358}), despite the presence of 5 µM $ZnCl_2$ in the protein buffer.

A similar representation (as complement of Figure 43) of inhibitors to the newly discovered exosite on proBFT-3, representing the chain B of the crystallographic asymmetric unit of proBFT-3, is illustrated in **Figure 44**.

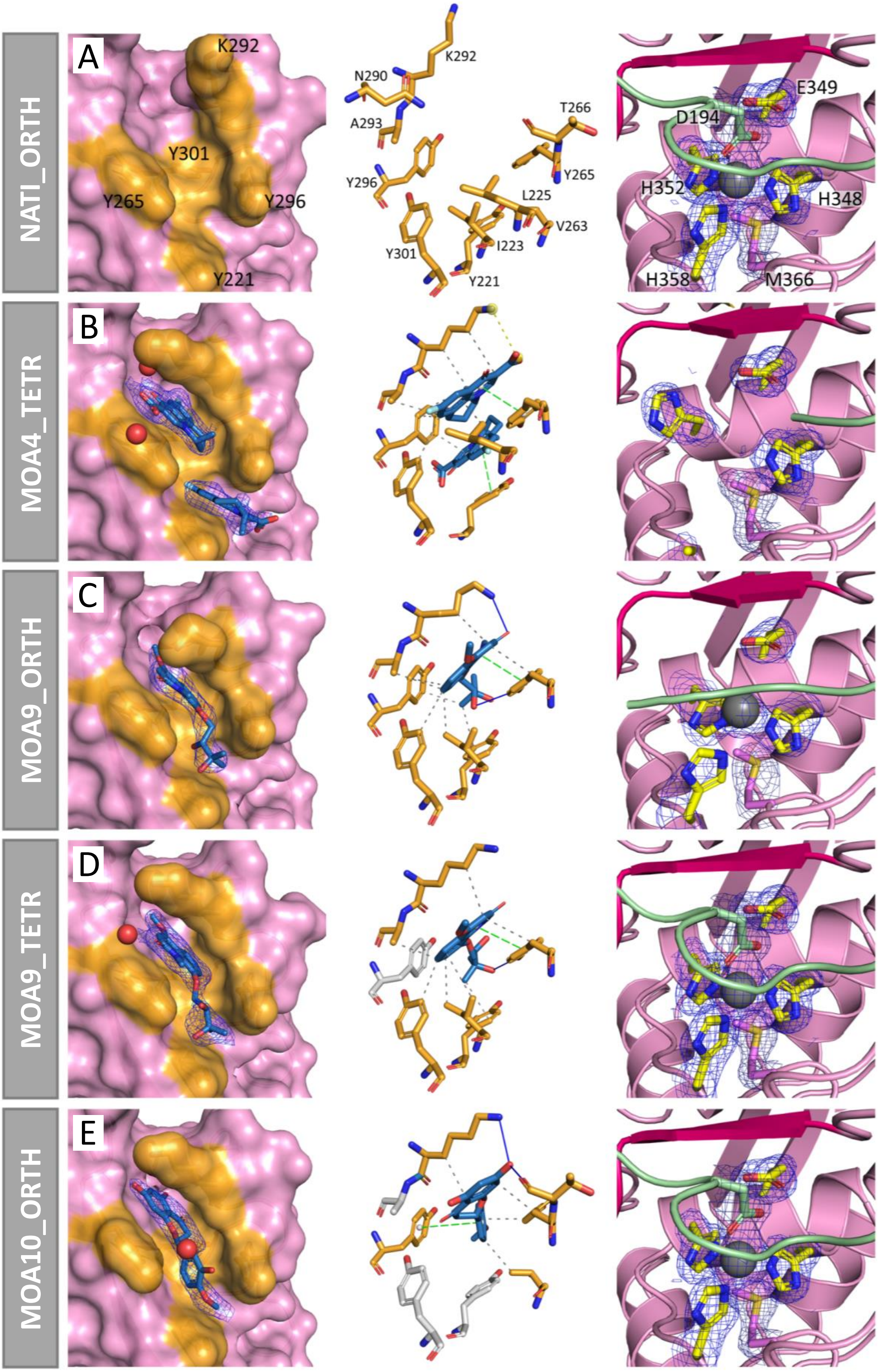

NATI_ORTH
A
K292
Y301
Y265
Y296
Y221
K292
N290
A293
T266
Y265
Y296
L225
V263
I223
Y301
Y221
E349
D194
H352
H348
H358
M366
MOA4_TETR
B
MOA9_ORTH
C
MOA9_TETR
D
MOA10_ORTH
E

Figure 44. Binding of inhibitors to the newly discovered exosite on proBFT-3 (representation of the chain B of the crystallographic asymmetric unit as complement of Figure 43). On the left, close-up views of the inhibitor-binding exosite are presented. In the middle, the protein-ligand interaction network that was identified is shown. On the right, the proteolytic active site is depicted. It is worth noting that for each of the structures, chain B of the crystallographic asymmetric unit was depicted, with the same orientations and settings as for chain A in the Figure 43.

Overall, the three MOA compounds were found to adopt a similar orientation and perform analogous interactions (**Figure 45**), except for MOA4, which displayed an extended binding interface due to the presence of the second bound molecule.

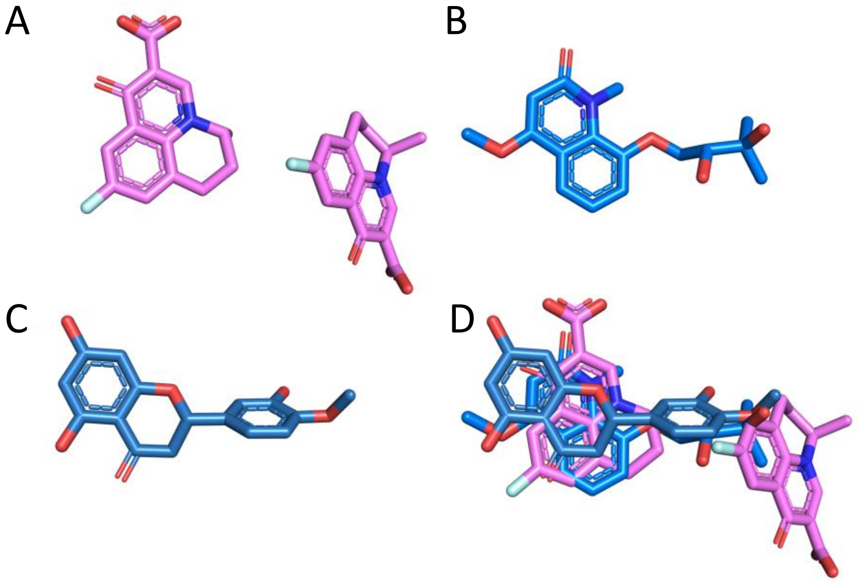


Figure 45. Molecular representation of the proBFT-3 exosite inhibitors. (A) MOA4, (B) MOA9 and (C) MOA10. (D) Superimposition of three inhibitory MOA molecules as found at the proBFT-3 exosite.

To assess the electron density and validation of the binding, ligand-omitted and bulk solvent-excluding mFo–DFc polder maps (215) were calculated using PHENIX (196) and contoured at 3.0 σ in Coot (216). These maps demonstrated the binding of all exosite-binding inhibitory compounds (**Figure 46**).

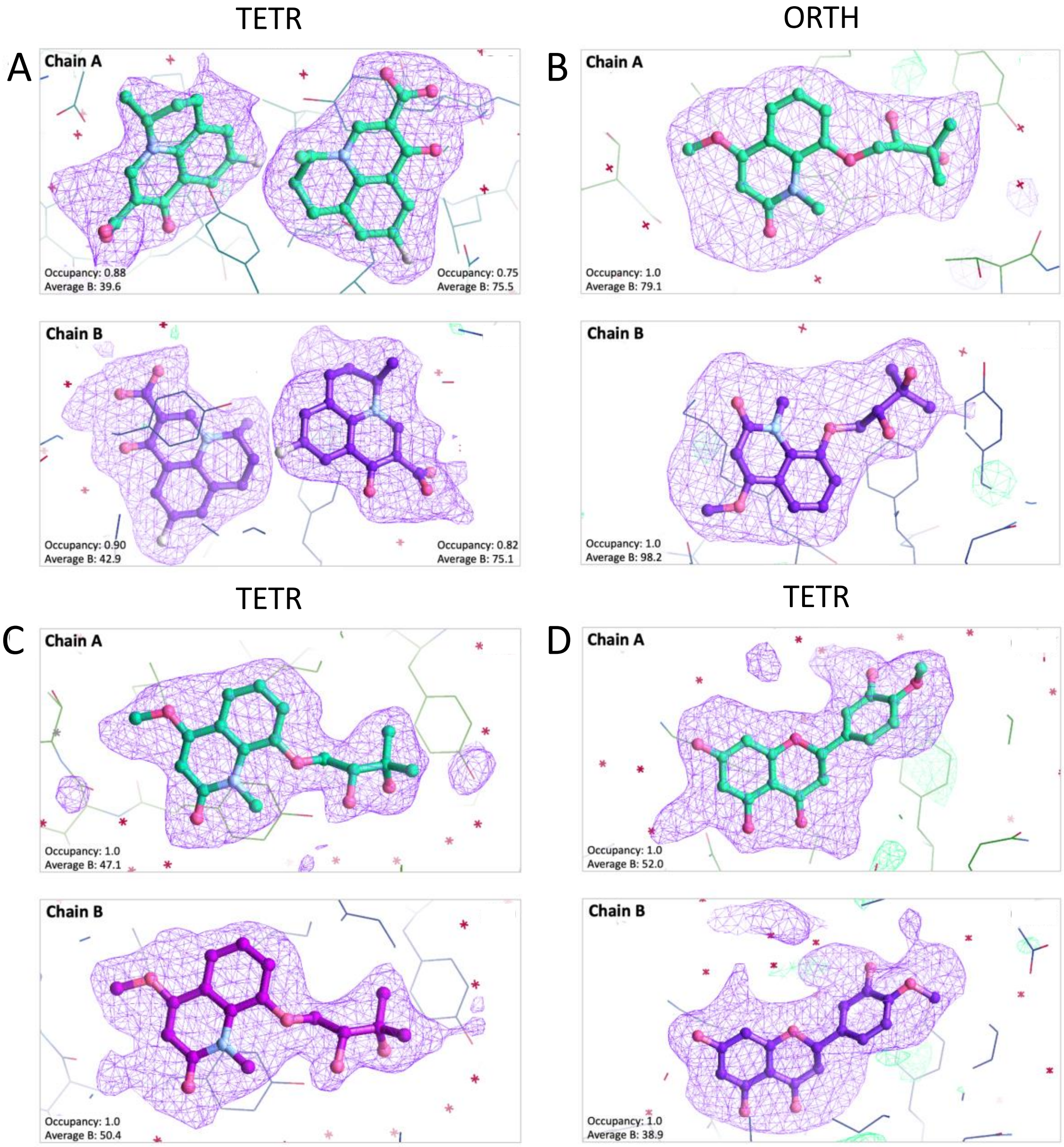

TETR
ORTH
A
Chain A
Occupancy: 0.88
Average B: 39.6
Occupancy: 0.75
Average B: 75.5
Chain B
Occupancy: 0.90
Average B: 42.9
Occupancy: 0.82
Average B: 75.1
B
Chain A
Occupancy: 1.0
Average B: 79.1
Chain B
Occupancy: 1.0
Average B: 98.2
TETR
TETR
C
Chain A
Occupancy: 1.0
Average B: 47.1
Chain B
Occupancy: 1.0
Average B: 50.4
D
Chain A
Occupancy: 1.0
Average B: 52.0
Chain B
Occupancy: 1.0
Average B: 38.9

Figure 46. Ligand-omitted and bulk solvent-excluding mFo-DFc electron density maps of the identified exosite inhibitors are depicted. Polder omit maps of were generated for (A) MOA4, (B) MOA9_ORTH, (C) MOA9_TETR, and (D) MOA10 using PHENIX. A solvent-exclusion radius of 3 Å and a resolution factor of 0.25 were applied. The electron density maps were contoured at 3.0 σ within Coot and are represented as a mesh around the respective exosite-binding ligand(s).

An interesting observation was that the binding of inhibitors at the exosite had an impact on the proBFT-3 protein crystallization. While the non-ligated proBFT-3 crystalized in the ORTH space group $P2_12_12_1$ and diffracted to 1.84 Å resolution, the complex structures with MOA4 (1.95 Å), MOA9 (1.84 Å), and MOA10 (2.03 Å) crystallized all in the TETR space group $P4_12_12$. For MOA9, it was possible to obtain an ORTH structure as well, but at a significantly lower resolution of 2.70 Å. Interestingly, the crystallographic asymmetric unit, which is the building block of both space groups, remained protein-wise identical. In both space groups, two protomers packed back-to-back with a protein-protein interface area of approximately 1,000 Å². According to the biocomputational analysis with Protein Interfaces, Surfaces, and Assemblies software (PISA) (202) of this surface indicated that this surface was not relevant for dimerization. However, experimental findings using size-exclusion chromatography (SEC) revealed a concentration-independent monomer-dimer equilibrium with a significant interface stability based on the immediate reinjection of the dimer peak (**Figure 47**). In the crystallization experiments, both the monomeric and dimeric peaks were concentrated together. It is important to note that since the interface was largely formed by the prodomains, it is unlikely for the activated BFT-3 protein to dimerize.

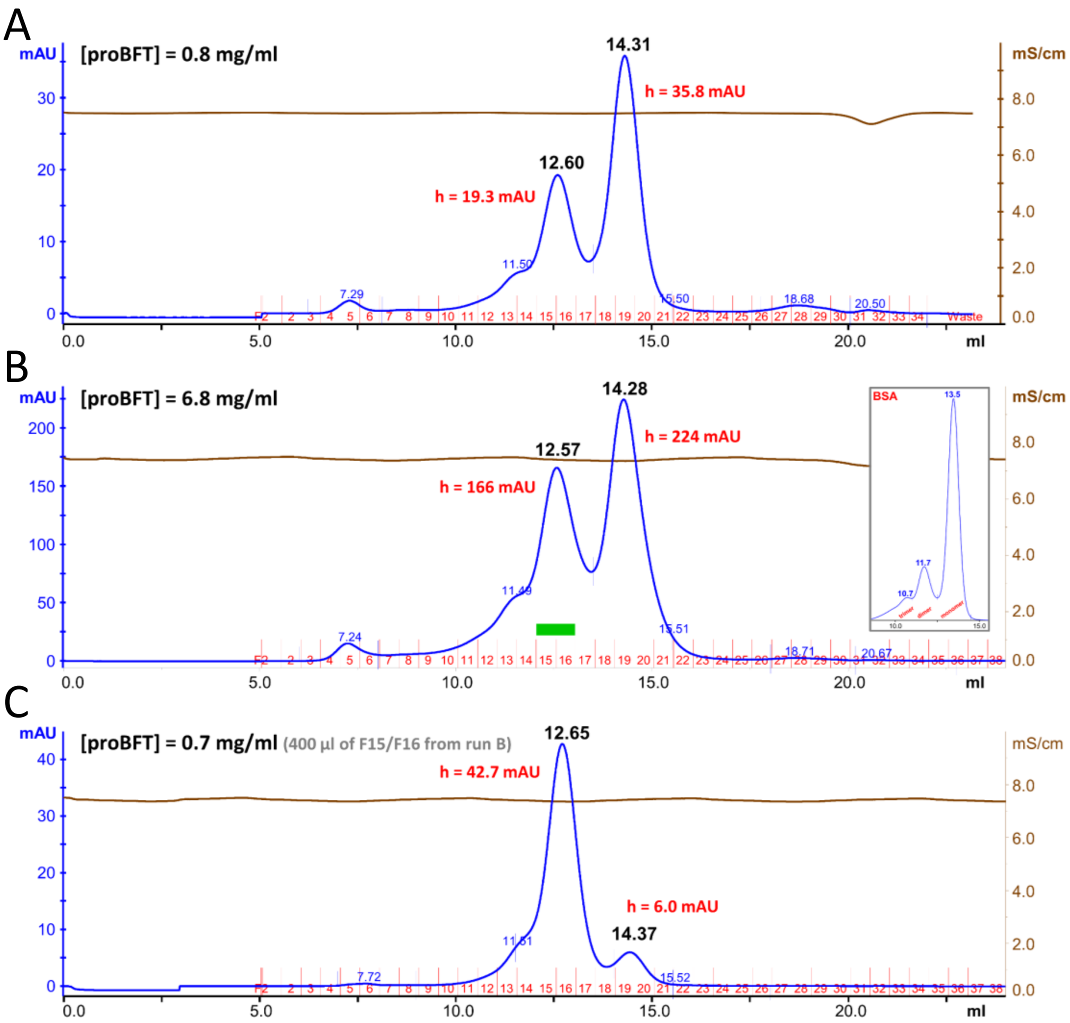


Figure 47. SEC was performed on proBFT-3 prior to protein crystallization experiments. The results showed that proBFT-3 existed as a dynamic monomer-dimer equilibrium at both low (A) and high (B) protein concentration. However, the dimer displayed a certain level of stability, as immediate (C) reinjection of the dimer peak samples (0.4 mL of combined fractions F15 and F16; signaled with a green bar) resulted in a predominantly dimeric size exclusion chromatogram.

Fractions containing both monomeric and dimeric proBFT-3 (F14 to F21) were pooled, concentrated, and subsequently used for the crystallization experiments. Based on a calibration run using bovine serum albumin (BSA) as a reference (shown as an insert in panel B), the apparent MW of monomeric proBFT-3 was determined to be 48 kDa, while that of the dimeric proBFT-3 was determined to be 95 kDa. The running buffer used for the SEC experiments consisted of 50 mM NaCl, 20 mM Tris-HCl, pH 8.0.

In addition, in the co-crystallization of proBFT-3 with the exosite inhibitors, crystals were observed to typically appear within a few days and reached their maximum size after approximately two weeks (**Figure 48**).

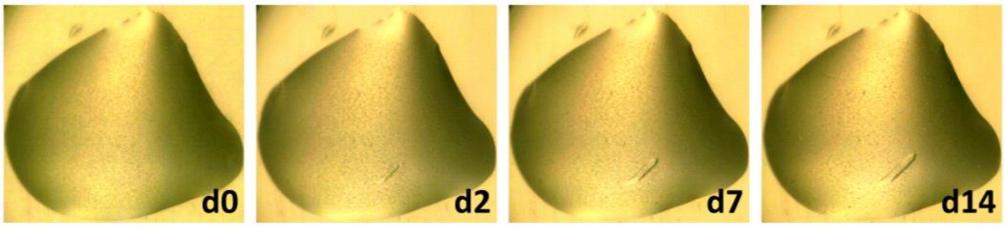


Figure 48. An illustration of crystal growth in the case of MOA10 in the sitting-drop vapor-diffusion format, utilizing 100 nL of protein and 100 nL of mother liquor, is provided as an example. The crystallization condition employed was as follows: 25% (w/v) PEG 3350, 0.2 M ammonium acetate, 0.1 M Bis-Tris HCl, pH 5.5. The temperature during the crystallization process was maintained at 4 °C. d0: day 0; d2: day 2; d7: day 7, and d14: day 14.

Of note, the binding of the inhibitors to the exosite deeply impacted the binding of the catalytic zinc despite the fact that the active site was approximately 25 Å apart. While the parameters of the thermal displacement of the active site residues were only slightly lower in the inhibitor complexes comparing to the overall structure (-12%), this effect was more than 2-fold larger in the case of non-ligated structure (-26%), which indicated that the binding of the inhibitors caused a destabilization effect in the active site. In a similar way, when analyzing the zinc ion of the active site, an increase in its parameters of thermal displacement upon inhibitor binding was observed, which was typically accompanied by a decreased in metal occupancy. Importantly, in the case of the MOA4 complex structure, this impact was clearly most pronounced, where the active site (especially the H^{358}) was distorted in both chains of proBFT-3. Thus, the zinc occupancy was reduced to 0.34 for the chain A, while the metal was completely missing in the chain B (**Figure 43 and Figure 44, right panels**). In the case of MOA9, a similar impact was observed in the ORTH space group, where the occupancy dropped to 0.22 for the chain A and 0.59 for the chain B. Compared to non-ligated proBFT-3, a 3.3-fold lower occupancy of the zinc ion (calculated as the zinc b-factor normalized to full-occupancy divided by the overall b-factor of the structure) was observed. Even for MOA10, where the full occupancy of zinc was observed, and for the MOA9 complex structure in the TETR space group, the stability of zinc was decreased in approximately 20%, despite observing a little effect at first sight. At this point, it is important to point out that all crystallization experiments were performed in the presence of 5 µM $ZnCl_2$, and, as a result, zinc occupancy in solution may be even lower. This aspect was consistent with the facts that the tested MOA compounds were identified against the zinc-free proBFT-3 (assays performed adding EDTA), their higher affinity toward the zinc-free conformational state, and the observed competitive interplay between the binding of compounds at the identified exosite and the occupancy of the zinc at the catalytic site.

To sum up, this newly discovered exosite played a crucial role in influencing the active site geometry and, consequently, the inactivation of proBFT-3 through a unique mechanism of allosteric inhibition. Furthermore, as the exosite was located within the CatD, a similar destabilizing and inhibitory effect was observed for activated BFT-3. Therefore, because the inhibitors did not bind into the active site, they allosterically affected both proBFT-3 and BFT-3.

4.3.6. MOA4 showed a greater antimicrobial effect *in vitro*, compared to MOA9 and MOA10 compounds

From the 1,120 chemical compounds tested in the experimental HTS, 3 of them (MOA4, MOA9 and MOA10) were the selected compounds that, by using several techniques, exhibited the most promising results by binding specifically to and showing an inhibitory effect against the BFT-3 toxin. The next step consisted on evaluating the effect of the identified compounds as potential candidates also in *in vitro* studies, but going one step further, by assessing the effect of these compounds in bacterial cultures of *B. fragilis*, which is the responsible for synthetizing and secreting this toxin.

It should be noted that the first step to do was to learn how to culture the pathogen *B. fragilis* in anaerobiosis, and to learn all the special requirements that these bacteria need in these particular conditions. Once these skills were acquired, the antimicrobial activity of the MOA4, MOA9 and MOA10 compounds was evaluated in both enterotoxigenic and nontoxigenic strains of *B. fragilis*. The purpose was to determine the MIC_{90} of these compounds, which indicated the compound concentration at which bacterial growth was inhibited by 90%.

Compounds MOA4 and MOA10 were found to inhibit bacterial growth better than MOA9. Besides, for these two compounds almost no differences were found between strains, only a slightly lower bacterial growth in the case of the nontoxigenic strain for MOA4. However, in addition to the fact that MOA9 was considerably less effective in inhibiting bacterial growth, it was observed that the nontoxigenic strain was more sensitive to this compound (**Figure 49**).

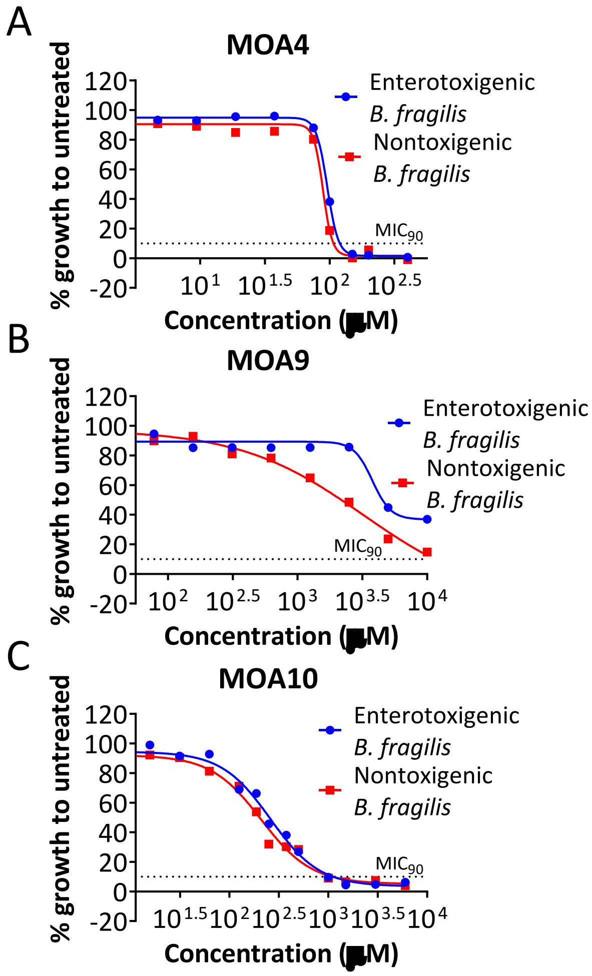


Figure 49. MIC$_{90}$ determination curves for (A) MOA4, (B) MOA9 and (C) MOA10 compounds on both strains of *B. fragilis*.

These results allowed the determination of the MIC$_{90}$ for MOA4 and MOA10 in both strains, but in case of the MOA9 only an estimation was performed (**Table 9**). For the following assays that will be explained below, the MIC$_{90}$ value will correspond to the 1X value of the compound used in those assays.

	***B. fragilis*: MIC$_{90}$ (µM)**	
	Nontoxigenic strain	**Enterotoxigenic strain**
MOA4	~110	~117
MOA9	>10,000	>10,000
MOA10	~1,000	~1,000

Table 9. MIC$_{90}$ determination for the MOA4, MOA9 and MOA10 compounds in *B. fragilis* cultures.

Then, to determine whether these compounds had a bactericidal or bacteriostatic effect, TKA assays were performed with MOA4 and MOA10 on both strains of *B. fragilis*. However, because these assays required a considerably high amount of each compound, the assays were not performed with MOA9, since it was much less available. Therefore, results obtained using MOA4 and MOA10 compounds are detailed below (**Figure 50 and Figure 51**).

In the case of MOA4, at the most active concentrations 4X and 10X, it inhibited bacterial growth up to $\sim 2.5\log_{10}$ on both strains, but regarding these data, they were not considered purely bactericidal (i.e., $< 3\log_{10}$-fold decrease in CFU). Besides, at a concentration of 1X MOA4 there was also a reduction in the bacterial growth, showing a bacteriostatic effect after 24 h, since the initial inoculum did not grow in that period of time, although later after 48 h there was a rebound and an increase of more than $1\log_{10}$ in CFU could be observed. In the case of the subinhibitory concentrations (below the previously stablished MIC_{90}), they had some activity since the growth of the inoculum slightly increased until 24 h, but from that moment there was a rebound that reached levels similar to those of the inoculum. Besides, it seemed that there were no differences in MOA4 activity between the two bacterial strains. Thus, MOA4 was a dose-dependent compound and, increasing both the dose and time, it changed its profile from a bacteriostatic compound to a quasi-bactericidal compound.

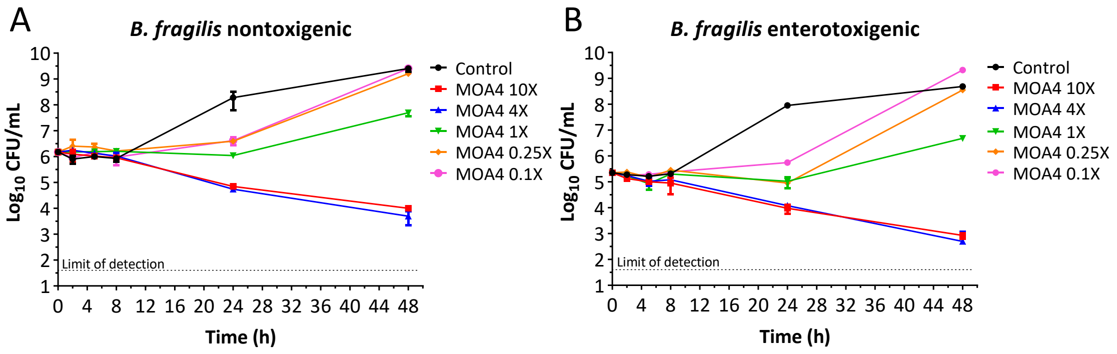


Figure 50. Time-kill kinetics assay in the antimicrobial characterization of MOA4 in *B. fragilis*: (A) nontoxigenic strain, (B) enterotoxigenic strain. In these assays, several concentrations of the MOA4 compound were tested for 48 h.

In the case of MOA10, it was observed that at 4X and 10X concentrations there was a turbidity change in the culture medium, which could be indicating that the compound slightly precipitated/aggregated. At 1X concentration, a bacteriostatic effect was achieved, as in the 10X concentration. However, the 4X concentration did not produce the expected inhibitory effect, possibly due to aggregation/precipitation of the compound, lowering the effective concentration in solution. In this assay, in general a decrease in bacterial growth up to 24 h was observed, but then under several conditions there was a bacterial growth rebound. In the case of the subinhibitory concentrations (0.25X and 0.1X), there was a slight decrease in growth with respect to the inoculum, above all up to 24 h. Taking into account the differences between strains, a slightly higher inhibitory effect was seen in the case of the enterotoxigenic one.

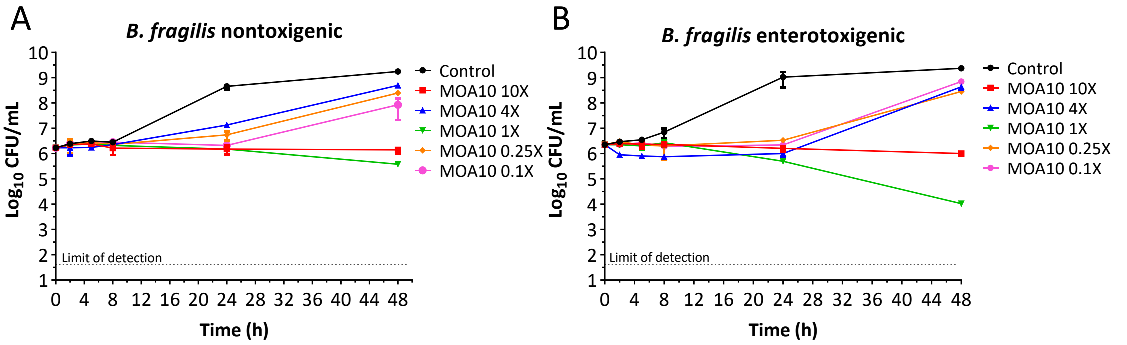


Figure 51. Time-kill kinetics assay in the antimicrobial characterization of MOA10 in *B. fragilis*: (A) nontoxigenic strain, (B) enterotoxigenic strain. In these assays, several concentrations of the MOA10 compound were tested for 48 h.

The results obtained in both MIC_{90} determination and TKA assays suggested that MOA4 compound had a greater antimicrobial effect *in vitro*, compared to MOA9 and MOA10 compounds, by more potently reducing bacterial growth of *B. fragilis* cultures.

4.3.7. MOA4 may have some synergy effect with MTZ, but not with FOX in *B. fragilis* cultures

In order to determine the bacterial sensitivity to some human clinical antibiotics of the *B. fragilis* strains of our research group, and compare to the average MIC_{90} calculated for the different existing strains (**Figure S1, Annex IV**), the antibacterial effect of some antibiotics clinically used in the treatment of infections produced by this bacterium in humans was evaluated (**Figure 52**). The clinical antibiotics selected for this purpose were MTZ and FOX.

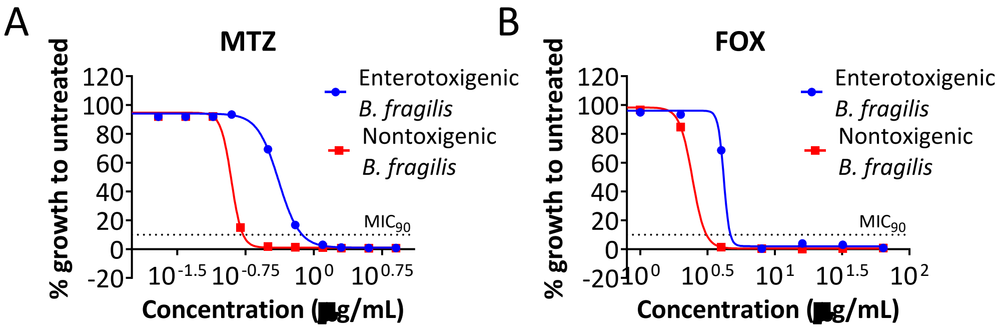


Figure 52. MIC_{90} determination curves for (A) MTZ and (B) FOX antibiotics on both strains of *B. fragilis* of our research group.

The MIC_{90} determination allowed a comparison of our *B. fragilis* strains with strains of this bacterium registered worldwide. Thanks to these experiments it was possible to verify that the strains from our laboratory had a sensitivity approximately within the expected range for the commercial antibiotics MTZ and FOX (**Table 10**).

	***B. fragilis*: MIC$_{90}$**	
	Literature	**Our strains**
MTZ	0.5-1 µg/mL (2.9-5.8 µM)	0.2-0.7 µg/mL (1.2-4.1 µM)
FOX	4-16 µg/mL (9.4-37.4 µM)	3.2-4.7 µg/mL (7.5-11 µM)

Table 10. Comparison of MIC_{90} data between all the registered *B. fragilis* strains and our *B. fragilis* strains. It should be noted that in human clinical practice it is not usual to refer to the concentration of a drug in molar units, but it is much more common to express it in mg/L or µg/mL. To compare this table with the values of the Figure X, here the results are expressed in µg/mL, but also in µM for better interpretation of the synergy assay that will be detailed below. MW MTZ: 171.16 g/mol; MW FOX: 427.45 g/mol.

Once the cytotoxic effects of two clinically used antibiotics (MTZ and FOX) were evaluated in both strains of *B. fragilis*, and data were compared to the MIC_{90} values calculated for all the identified strains all over the world, synergy studies of these two antibiotics with MOA4 compound were performed. Synergy studies between compounds are highly interesting in the drug discovery process, since the doses currently used for certain compounds already used in human clinical practice can be lowered if new compounds presenting synergies with them are identified.

Results showed that when MOA4 compound was combined with FOX at different concentrations, no synergistic effects were observed. However, the combination of MTZ with MOA4 compound showed a reduction in the bacterial growth (**Figure 53**).

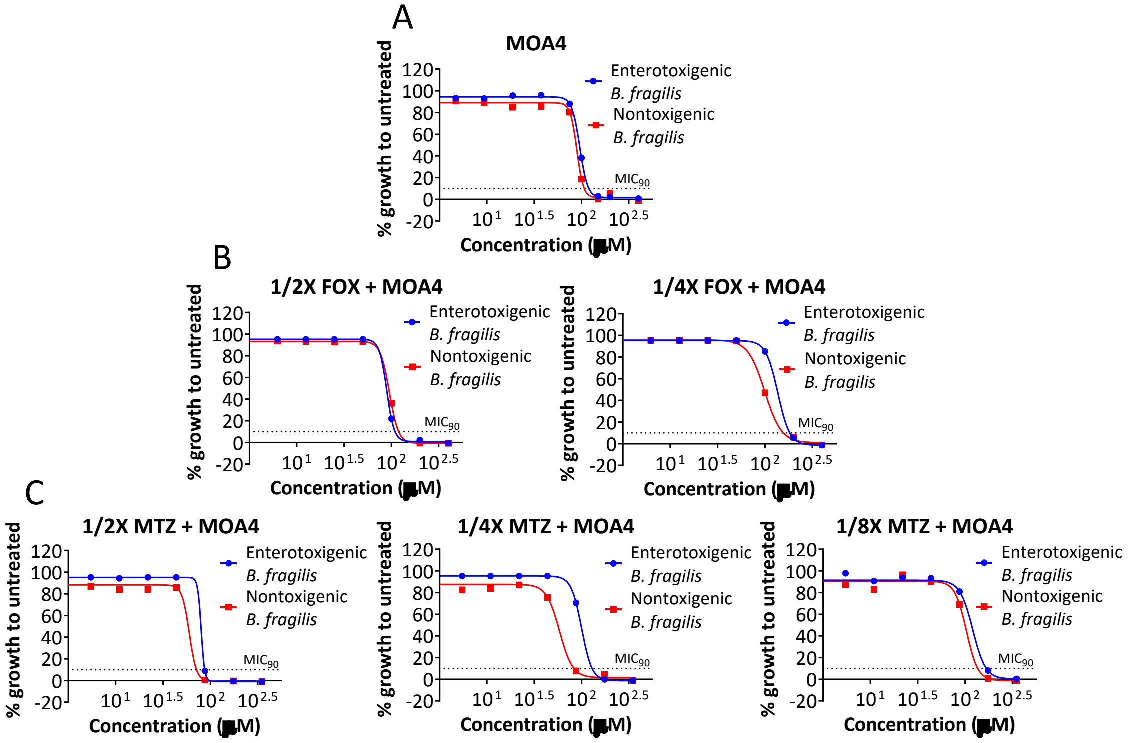


Figure 53. Analysis of possible synergy effects between (B) FOX and (C) MTZ (both PC at fixed concentration calculated from their MIC_{90} (1X)) and MOA4 compound (SC at different concentrations by serial dilutions). (A) As a control for the synergy assays, the effect of MOA4 was represented. Data depicted is mean of two independent experiments, and SD is <10%.

In these synergy studies, the MIC_{90} for all the tested combinations (**Table 11**) and the evaluation of the existence of synergy between MOA4 and MTZ (**Table 12**) were determined. Of note, slightly higher than expected MIC_{90} values, compared to the control (MOA4 alone), were obtained in two cases: the 1/4X FOX + MOA4 and 1/8X MTZ + MOA4. These assays did not result in a fixed MIC_{90} value, but actually ranges of MIC_{90} values that may vary somewhat between each assay and plate, but these differences were not significant within the concentration order in which they were found.

	Synergy studies in *B. fragilis*: MIC_{90} (µM)	
	Nontoxigenic strain	**Enterotoxigenic strain**
MOA4	110-115	117-120
1/2X FOX + MOA4	117-120	110-115
1/4X FOX + MOA4	143-150	160-180
1/2X MTZ + MOA4	71-73	85-87
1/4X MTZ + MOA4	80-82	120-127
1/8X MTZ + MOA4	125-135	145-165

Table 11. MIC_{90} determination in the synergy studies for the MOA4 in combination with MTZ and FOX in both *B. fragilis* strains. MIC_{90} is represented as a range from the analysis of three assays.

Condition	**Analysis of synergy studies in *B. fragilis*: MIC_{90} (µM)**				
	MIC_{90} PC	**MIC_{90} SC**	**MIC_{90} (PC+SC)**	**MIC_{90} (PC) + MIC_{90} (SC)**	**MIC_{90} (PC+SC) ≤ 0.25 [MIC_{90} (PC) + MIC_{90} (SC)]**
1/2X MTZ + MOA4 Nontoxigenic strain	1.2	115	71	116.2	No
1/2X MTZ + MOA4 Enterotoxigenic strain	4.1	120	85	124.1	No
1/4X MTZ + MOA4 Nontoxigenic strain	1.2	115	80	116.2	No

Table 12. Evaluation of the existence of synergy between MOA4 and MTZ. Of note, only data where possible synergy could exist were analyzed. The remaining conditions, where it was clear that there was no synergy, are not represented.

These results indicated that the combination of MTZ with MOA4 showed higher antibacterial effect than both compounds alone. However, as the MIC_{90} of the SC compound (MOA4) was not reduced ≥4-fold in the presence of the PC compound (MTZ), compared to the MIC_{90} of the SC compound alone, a purely synergy effect was not achieved. These results suggest the possibility of further synergy studies to better evaluate the possible synergistic effect between MOA4 and MTZ.

4.3.8. First steps, establishment of standard conditions and aims in the implementation of *G. mellonella* as a new infection animal model for *B. fragilis*

Once the MOA4, MOA9 and MOA10 compounds were selected based on the results obtained in different *in vitro* techniques, *in vivo* studies were needed to continue the process of compound identification and development in the drug discovery.

As previously mentioned in the introduction, *G. mellonella* larvae are increasingly used in the field of microbiology and drug discovery for the past few years. However, to date the use of this animal model to evaluate the infections produced by *B. fragilis* has not been described, nor a searching process for new drugs using it has been developed. Therefore, one of the most ambitious aims of this Doctoral Thesis was to implement and develop a new infection animal model for the study of this pathogen, i.e., *G. mellonella* as an infection animal model for *B. fragilis*, in order to evaluate both the evolution of the infection process and the effect of the identified candidate compounds (MOA4, MOA9 and MOA10) *in vivo*.

Hence, before infecting *G. mellonella* with *B. fragilis*, it was necessary to acquire the necessary knowledge for working with these worms, as well as to develop and learn the suitable conditions for handling them. The project began with the study of larval management after their delivery, suitable environmental conditions for their maintenance, and the recognition of the ideal size and appearance of the worms for the intended studies (**Figure 54**). It was also important to evaluate the appropriate injection volumes and number of worms (n) to be used in each assay. All these considerations were taken into account based on studies with other pathogens in *G. mellonella*, performing the appropriate modifications in the protocols.

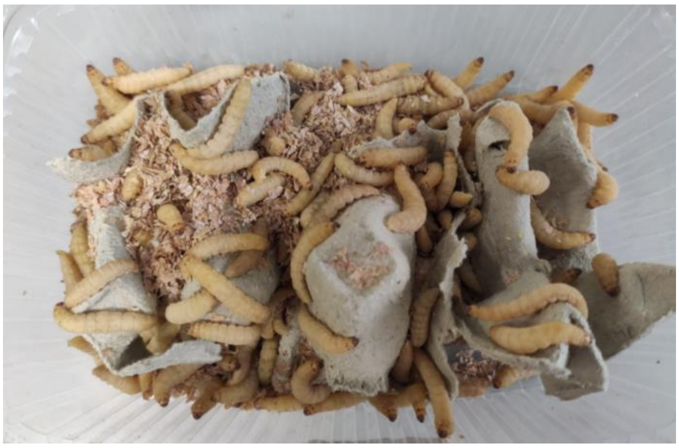


Figure 54. *G. mellonella* worms after the delivery. Larvae with approximate a weight of 180-250 mg, mobile and with a slightly yellow/tan coloration were classified after the delivery as larvae in good health conditions for *in vivo* assays. There should be a very small number of larvae with black spots or discolorations along the body, and it was necessary to work with worms of the same size and around 2 cm, which were between the 5th and 6th stage of development.

Once learned how to manipulate *G. mellonella* larvae, and before proceeding to detail the results obtained, it is important to explain the objectives that were then established, for a better understanding of the assays. In particular, three main objectives were defined:

- First, to analyze the necessary CFU/mL to induce a clear and rapid bacterial infection in *G. mellonella*. For this purpose, different dilutions of bacteria were tested for calculating the cytotoxicity curves.
- Second, to know what concentrations of MOA compounds the larvae were able to tolerate. As in the previous objective, cytotoxicity curves were performed using different concentrations of compounds.
- Third, to evaluate whether after infecting the larvae with both strains of *B. fragilis*, and a subsequent treatment with compounds, a reduction in the infection of the treated worms, compared to the untreated ones, was observed.

Taking into account these three objectives and the recommendations found in the literature for other pathogens, it was established a first injection of 10 µL/worm to produce the infection with *B. fragilis*, as well as a second injection of 10 µL/worm to perform the treatment with the compounds. Therefore, for infected and treated worms, two injections of 10 µL were finally needed, i.e., 20 µL total volume injected. For this reason, in order to get a suitable comparison of the three cases mentioned in the previous objectives, it was necessary to always perform two injections to each worm, regardless of the sought goal. Hence, in the case of studying only the bacterial concentration necessary to induce an infection, an initial injection with the bacteria was performed and then, a mock second injection with 1X PBS simulating a "treatment" was needed. In the same way, when evaluating what dose of compound the larvae tolerated, an initial mock injection simulating an "infection" was performed, also with 1X PBS, and then a second injection with the real treatment.

In relation to handling and learning how to handle larvae, initially n=4-7 worms were used for each assay. This is an important aspect to note, since enough reproducibility was only obtained once the number was increased to n=10. In this way, we learned to handle the worms and to perform injections in them with fluency, with injections of 10 µL of 1X PBS in 10 worms/Petri dish, until human manipulation did not affect the health status of the worms.

4.3.9. MOA4 was the less toxic compound in *G. mellonella* larvae, followed by MOA9 and MOA10

Once the necessary experience in working with *G. mellonella* larvae was acquired, the safety range of the compounds identified as potential BFT-3 inhibitors was evaluated in non-infected worms. Performing these assays, it was possible to know what concentrations of the selected compounds were tolerated by the larvae and to analyze what cytotoxic effect these compounds exerted.

In order to analyze the intrinsic cytotoxic effect of the selected compounds in *G. mellonella* larvae, 10 µL injections of 1X PBS to simulate a bacterial infection, followed by 10 µL injections of different concentrations of MOA4, MOA9 and MOA10 were performed into the worms. **Figures S2, S3, S4 and S5** summarize the 24 h monitoring of the different conditions tested with the worms (**Annex IV**).

In the case of MOA4 (**Figure 55**), at the highest concentration tested (1,200 µM) a clear toxic effect on the worms was observed at approximately 24 h. At lower concentrations (≤ 400 µM), the worms maintained a survival rate of 60-80%, similar to that present in the controls, which were treated with no compound. Then, the MOA4 compound seemed to be fairly no toxic at concentrations below 400 µM.

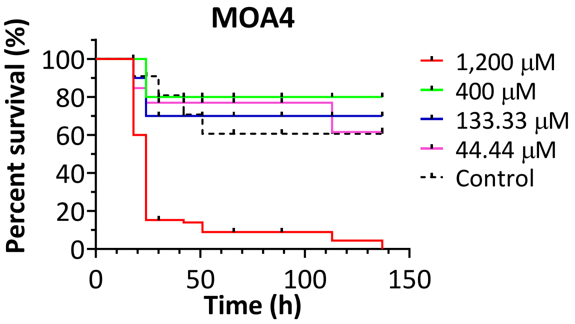


Figure 55. Effect on survival rates of different concentrations of MOA4 compound on *G. mellonella* larvae. Survival analysis rates were determined using the Kaplan-Meier estimator and the data came from the means of two independent experiments.

In the case of compound MOA9 (**Figure 56**), at the highest concentration of 1,200 µM a reduction in the survival rate to 30% at 48 h approximately was observed. At lower concentrations (around 400 µM), the viability of the worms was reduced to about 40% at 24 h. The concentration of 133.33 µM reduced survival only to 70% at 24 h, similar to that obtained in the controls.

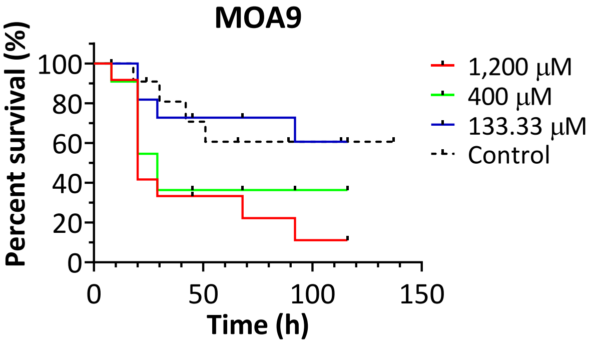


Figure 56. Effect on survival rates of different concentrations of MOA9 compound on *G. mellonella* larvae. Survival analysis rates were determined using the Kaplan-Meier estimator and the data came from the means of two independent experiments.

Finally, in the case of the MOA10 compound (**Figure 57**), a higher toxicity was observed at the concentrations tested in comparison to the MOA4 and MOA9 compounds.

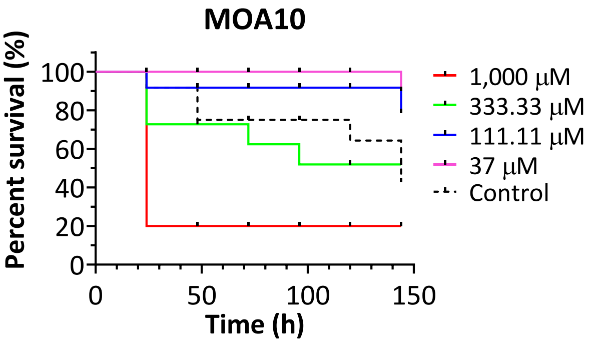


Figure 57. Effect on survival rates of different concentrations of MOA10 compound on *G. mellonella* larvae. Survival analysis rates were determined using the Kaplan-Meier estimator and the data came from the means of two independent experiments.

With these assays it was possible to obtain information about the intrinsic effect of the compounds and also about the non-toxic (or low toxicity) concentrations of compound to be used in worms.

4.3.10. Success in developing a new *G. mellonella* animal infection model to study *B. fragilis* infections

First, it was necessary to develop an infection model for *B. fragilis* based on *G. mellonella.* As previously mentioned, these larvae are widely used in the field of microbiology and drug discovery. However, the use of this animal model to evaluate the infection produced by *B. fragilis* has not been described yet, nor there is a stablished process for searching for new drugs against *B. fragilis* using *G. mellonella*.

The infection process, defined as a condition in which bacteria causing diseases have entered the body, can be easily observed in these worms by the process of melanization. In this process, worms change from a yellowish/tan coloration to showing a black spots pattern along the body and, after a while, a totally blackish color once the infection has severely progressed.

In this assay, the minimal CFU/mL of *B. fragilis* necessary to produce a clear and rapid infection in *G. mellonella* larvae was determined. For this purpose, a 10 μL injection was performed at different bacterial dilutions ($3\cdot10^9/2\cdot10^8/10^8/10^4/0$ CFU/mL), after which an additional 10 μL 1X PBS injection was carried out in each worm to simulate a treatment. **Figures S6, S7, S8 and S9** summarize the 24 h monitoring of the different conditions tested on the worms (**Annex IV**).

At the maximum concentration of *B. fragilis* tested ($3\cdot10^9$ CFU/mL), similar results were observed in worms treated with the nontoxigenic and enterotoxigenic strains. **Figure 58** shows a 40%

reduction in the survival rate of the worms at 48 h, after which survival drops to 20-10% over the next 48 h. When lower concentrations of bacteria ($2\cdot10^8/10^8/10^4$ CFU/mL) were inoculated, a reduction in survival rates were also seen with both bacterial strains after 48 h, but smaller than that observed at $3\cdot10^9$ CFU/mL. In the case of controls (worms not infected with bacteria, but inoculated with the same volume of 1X PBS), no reduction in survival rates were observed over the time period analyzed (with survival rates around 90% at the end of the assay).

The results of infection of *G. mellonella* with *B. fragilis* showed that at a concentration of $3\cdot10^9$ CFU/mL, 80-90% of the infected worms were dead within 72 h. Thus, it was observed that the higher CFU/mL of bacteria of both the nontoxigenic and enterotoxigenic strains, the lower the survival of the worms with respect to the controls, so that there was a dose-dependent effect in the infection process, thus confirming the successful development of *G. mellonella* animal infection model to study *B. fragilis* infections.

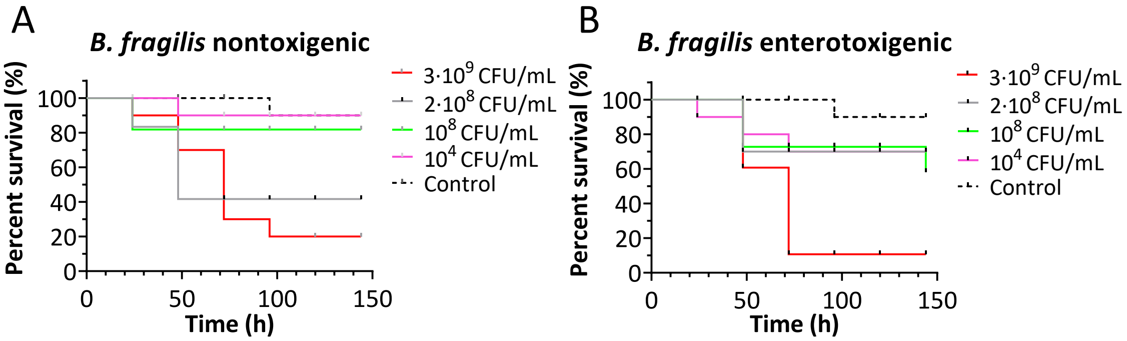


Figure 58. Effect on survival rates of different doses of (A) nontoxigenic and (B) enterotoxigenic *B. fragilis* culture injections on *G. mellonella* larvae. Survival analysis rates were determined using the Kaplan-Meier estimator and the data came from the means of three independent experiments.

4.3.11. The treatment of infected *G. mellonella* larvae with MOA compounds increased their larval survival rate

Finally, the antibacterial effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* worms infected with both strains of *B. fragilis* was tested. The goal was to assess whether the compounds were able to reduce the bacterial infection produced in infected worms after the compound injection treatment, compared to infected control worms injected with 1X PBS as mock treatment.

For this aim, an initial 10 µL injection of 10^9 CFU/mL of *B. fragilis* were performed in worms, followed by a second 10 µL injection of the MOA4, MOA9 and MOA10 compounds at a fixed concentration (200 µM). **Figures S10, S11, S12 and S13** summarize the 24 h monitoring of the different conditions tested on the worms (**Annex IV**).

As it can be observed (**Figure 59**), in the case of 10^9 CFU/mL *B. fragilis* infected worms and later treated with 200 µM MOA4 compound, they maintained survival rates around 80% at 122 h, while the untreated infected worms (controls) achieved only 20% survival rates after 24–48 h with a rapid onset of melanization. Therefore, MOA4 increased 4-fold larval survival (from 20% to 80%). In the case of this compound, similar results were obtained when worms were infected with the nontoxigenic strain, but the survival rate was slightly higher. In the case of MOA9, differences were observed between strains, since in the nontoxigenic strain there was a survival increase compared to the control group (MOA9 treated worms had a 30% survival increase after 24 h and 10% survival increase after 96 h), but in the enterotoxigenic strain a higher survival rate of the MOA9 treated worms was observed (30% from 24 h to 144 h). Therefore, this compound seems to be strain-dependent. For the MOA10 compound, there was an increase in larval survival (30-40%), but it reached similar values to those of the control group after 144 h.

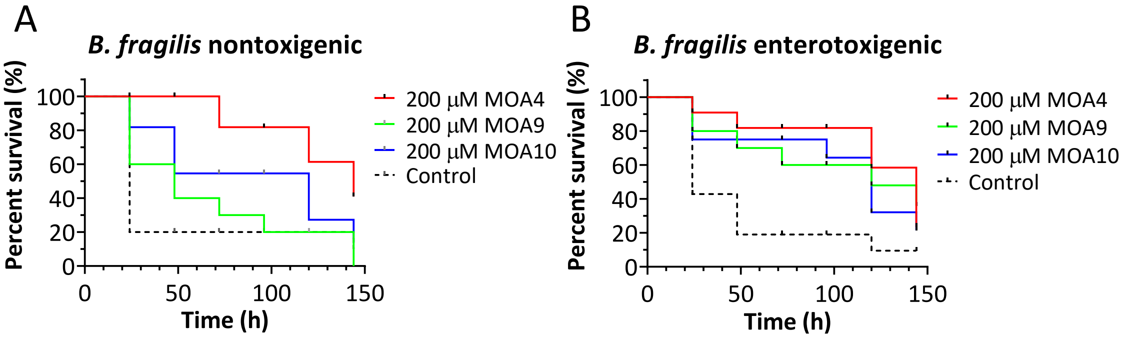


Figure 59. Effect on survival rates of 200 µM MOA4, MOA9 and MOA10 compounds on *G. mellonella* larvae after the infection with 10^9 CFU/mL of (A) nontoxigenic and (B) enterotoxigenic *B. fragilis* strains. Survival analysis rates were determined using the Kaplan-Meier estimator and the data came from the means of two independent experiments.

Because MOA4 compound was less toxic in worms, and it caused higher inhibition in bacterial cultures growth, it was tested at a lower concentration (100 µM) in worms (**Figure 60**), to evaluate whether similar survival rates were observed at a lower dose. After the 100 µM MOA4 treatment and 10^7 CFU/mL *B. fragilis* infection, the compound treatment increased by 40% larval survival in the nontoxigenic strain assay, while it increased by 70% larval survival in the enterotoxigenic strain.

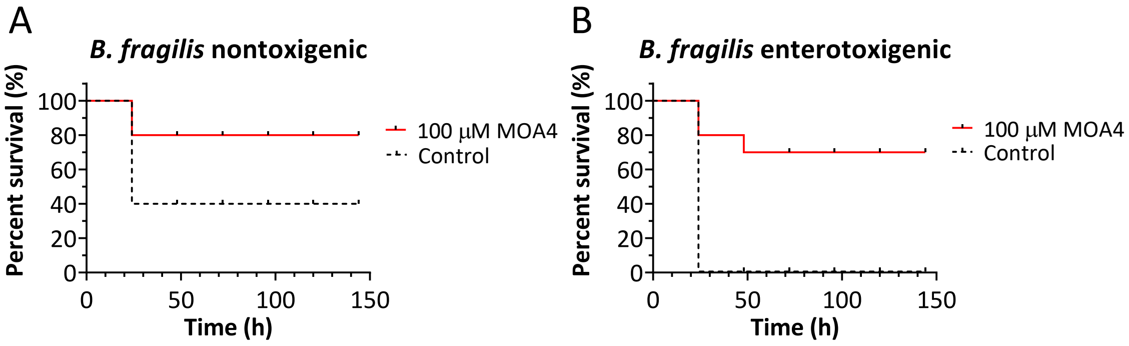


Figure 60. Effect on survival rates of 100 µM MOA4 compound on *G. mellonella* larvae after the infection with 10^7 CFU/mL of (A) nontoxigenic and (B) enterotoxigenic *B. fragilis* strains. Survival analysis rates were determined using the Kaplan-Meier estimator.

In brief, MOA4 compound was the most effective compound at lower concentrations *in vivo* against *B. fragilis* infections in *G. mellonella* worms, increasing the survival rates of *B. fragilis*-infected larvae compared to infected controls.

5. DISCUSSION

CRC is one of the most important gastrointestinal pathologies, being the 3rd most common type of cancer diagnosed, and being associated with a large number of environmental and genetic risk factors (7,8). One of these environmental factors is the presence of infectious agents, such as bacteria, either in disproportionate amounts or in inappropriate locations for them. Among these bacteria, *B. fragilis* is one of the most relevant clinical pathogens, implicated in the development of numerous infections, leading to the formation of intestinal abscesses, bacteremia, inflammatory diarrhea and intestinal inflammation (31,36–38), aspects that can potentially be implicated in the development of CRC (43,44,50,51).

AMR is a global health emergency mainly due to the continuous increase in the demand for antibiotics across many sectors, the natural selection and the irresponsible overuse of them (93). In fact, the WHO has long been warning about the enormous problem of multidrug-resistant bacteria, which are causing (and will cause) more deaths worldwide. Specifically, it is predicted that we are going to be in a post-antibiotic era that will result in more frequent infections and even small injuries might result in death if we fail to act against antibiotic resistance (92). Moreover, the situation is expected to worsen in the long run if we do not act correctly. It is estimated that up to 10 million people could die annually from AMR by 2050, according to studies by The Review on Antimicrobial Resistance, published in 2016 (96,97). Therefore, in order to reduce AMR, making informed and location-specific policy decisions regarding control programs to prevent infections, the access to essential antibiotics, and the research for the development of new antibiotics as well as vaccines is crucial.

In the context of AMR, *Bacteroides* spp. presents “the most antibiotic-resistant mechanism and the highest resistant rates of all anaerobic pathogens" (31). Then, a suitable use of effective treatments is essential for a favorable clinical outcome in infections caused by *B. fragilis*. This bacterium is commonly exposed to different antimicrobial and antibiotics, and as in other cases, *B. fragilis* evolved from being susceptible to becoming resistant to a broad spectrum of antimicrobial agents (42,67).

There is an important consideration to be done: while administering a broad-spectrum antibiotic may tackle the *B. fragilis* infection, it may affect as well the beneficial flora. In this way, orally administered broad-spectrum antibiotics may remove enteropathogens from the gastrointestinal tract, but they may also affect the commensal flora that human need. Thus, opportunistic microorganisms can colonize the intestine in the absence of this beneficial flora, and lead to dysbiosis and severe gastrointestinal diseases. That is why, there is an urgent necessity to develop specific antimicrobial treatments or to rethink the existing ones, in order to achieve a specific inhibition of BFT (the specific toxin from *B. fragilis*), i.e., to combat ETBF-mediated pathogenicity without disturbing the commensal microbiota (38,98,108–110,100–107). Accordingly, there is an indisputable need for better understanding ETBF targets, such as BFT-3, in order to design highly specific antimicrobials to tackle them (42,67,111).

The drug discovery against a target protein always benefits from a comprehensive structural and functional characterization. For this purpose, first of all, high quantities of the protein of interest, i.e., BFT-3, must be obtained. As expression system, *E. coli* strains were used, followed by a purification system, specifically using cobalt affinity columns. Once the pure protein was obtained in large quantities, and after verifying the correct identity of this protein by MALDI-TOF/TOF, the planned assays were performed.

Protein-bound zinc ions can be classified based on their functional roles as either structural, contributing to protein folding stability, or catalytic, participating in enzymatic processes. The structural protein-bound zinc ions are typically bound distal to the active site, often coordinated by cysteine residues, such as in the case of the HCV NS3 protease (181). On the other hand, the catalytic protein-bound zinc ions are predominantly complexed by aspartates, glutamates and histidines located at the active center, such as in the case of H348, H352, and H358 in BFT-3. In certain cases, such as in these three histidines of BFT-3, these amino acids contribute to both catalysis and overall protein stability (67). Despite the close relation between these roles, it is important to note that EDTA behaves differently depending on the protein. For instance, due to differences in the respective coordination spheres, EDTA effectively removes zinc from BFT-3 at pH 7, while EDTA fails to do so in HCV NS3 at pH >6 (217), despite the fact that EDTA exhibits a strong binding affinity for zinc in the picomolar range (218,219).

As previously mentioned, the protein is synthesized as immature protein (proenzyme or zymogen) or proBFT-3, and the active form is BFT-3, therefore, depending on the assay, it was of interest to have the protein in the immature or mature form, whose activation was performed with trypsin. Furthermore, based on the aforementioned, the assays were performed in the presence or absence of zinc by adding EDTA, and with or without previous trypsin treatment.

Several techniques were employed for the biophysical characterization of the protein, using the protein both with and without zinc, which verified that the protein is zinc-dependent. Significantly, a substantial unfolding in proBFT-3, although to a lesser extent than in NS3 (181), was observed when zinc was removed. Thus, moderate alterations in both the secondary structure content observed in the far UV-CD spectrum, related to the α-helical content, characterized by an unexpected increase in ellipticity, was detected. The near-UV CD spectrum indicated that there were small changes upon zinc removal in proBFT-3, therefore indicating that proBFT-3 remained, at least partially, folded upon the elimination of zinc. On the other hand, the intrinsic tryptophan fluorescence was reduced 10% in the absence of zinc, pinpointing conformational changes in the local microenvironment of at least one of the four tryptophan residues in the CatD of proBFT-3. One possible explanation for these findings was that zinc functions as an anchor between the upper and lower segments of the CatD of proBFT-3. Consequently, the absence of zinc increased the overall flexibility of the protein while minimally affecting the secondary structure content and the tryptophan environment. Certainly, when zinc was absent, there was a significant increase in the exposed surface area of the proBFT-3 to the solvent, as shown by the increment in ANS fluorescence quantum yield, indicating that some hydrophobic residues that were initially buried inside the protein, became solvent-exposed and bound to the ANS dye after removing the zinc. However, both the structural stability and the folding cooperativity decreased, as evidenced by the notably reduced thermal unfolding stability parameters: lower T_m and unfolding enthalpy through DSF assays. Since the presence of zinc bound to proBFT-3 significantly enhanced its thermal stability, these results further validated the key structural role of the catalytic zinc ion in proBFT-3, therefore indicating that the zinc binding conferred a considerable increase in the resistance of proBFT-3 to thermal denaturation. In fact, according to the thermal stability parameters obtained in DSF, it can be signaled that the interaction with zinc contributed to an estimated stabilization Gibbs energy of ~3.6 kcal/mol. This value significantly surpassed the intrinsic stabilization energy of the zinc-free state (~2.5-3.0 kcal/mol), and the zinc interaction accounts for ~60% of the total stabilization energy observed in proBFT-3.

All these assays together indicated that the conformation of the protein was different in the presence and the absence of zinc, confirming the importance of this metal cofactor in the folding of proBFT-3. Therefore, it was proposed that BFT-3 could be classified as a conditionally disordered protein (CDP), as previously established in the case of HCV NS3 (220). The way in which CDPs function are by temporarily adopting a well-folded structure depending on intrinsic or extrinsic environmental factors, while also experiencing situations in which they are partially or completely disordered. This conformational switch is as a regulatory mechanism that can be induced by factors such as post-translational modifications (PTMs), interactions with biological partners or changes in environmental conditions, such as alterations in zinc levels. It is noteworthy that CDPs generally possess low structural stability and share similar characteristics to intrinsically disordered proteins (IDPs), as they populate conformational states separated by small conformational energy gaps and rather easily interconverting. In the case of proBFT-3, it appears that the solvent and the protein alone are insufficient to drive the folding toward the fully structured state. However, upon zinc binding, the required interactions and structural context are available, enabling the folding process.

Importantly, there are several factors that contribute to the proposition that the partially unfolded state of proBFT-3 in the absence of zinc holds physiological significance: i) firstly, the tight spatial, thermodynamic and kinetic regulation of the intracellular zinc pool at subnanomolar concentrations (221–224); ii) the required proBFT-3 plasticity for secretion into the intestinal lumen, where the protein needs to adopt a sufficiently relaxed conformation for membrane translocation followed by subsequent refolding (similar to Tim15, a zinc-dependent protein that translocates to the mitochondria (225)); and iii) the rather low intrinsic (i.e., zinc-free protein) conformational stabilization Gibbs energy of proBFT-3. These facts could add an additional layer of proBFT-3 activity regulation alongside the well-known inhibitory role of the PD (76). While the partially unfolded state due to insufficient zinc would ensure minimal intracellular proBFT-3 activity, its structural plasticity could facilitate both the secretion and activation of the protein, the latter likely through the action of a host protease such as trypsin, an abundant protease in the natural niche of *B. fragilis*. Nevertheless, the potential for enhanced and sustained BFT-3 activity would be attainable upon zinc binding and subsequent stabilization of the protein within the gut lumen.

Given the plausible physiological relevance of the zinc-free state of proBFT-3, a search for small molecules capable of binding and stabilizing this partially disordered conformation was carried out in this Doctoral Thesis. Therefore, the aim was to trap the protein in an inactive state, achieved through an allosteric inhibition mechanism where the binding of a ligand modulates the conformational equilibrium between the active and inactive states (226,227). The resulting inhibitory mechanism could be categorized as competitive or mixed inhibition, depending on whether the substrate and inhibitor were mutually exclusive (i.e., the substrate and inhibitor solely interact with the folded and partially unfolded protein, respectively) or if a ternary complex could form, displaying reduced but detectable activity. In fact, it can be considered that inhibitor and zinc cofactor, although they would bind preferentially to different conformations, those conformations are in equilibrium, and both ligands would “compete” for binding to the protein. In this context, a HTS based on TSA was performed using the Prestwick Chemical Library, which was carried out under experimental conditions favoring the dissociation of zinc (by utilizing the chelating agent EDTA). Once all the thermal denaturation curves were analyzed, the compounds in which the curve shifted towards higher temperature values were selected. Thus, 11 known FDA-approved drugs that could potentially act as ligands for BFT-3 were identified. In this approach, a semi-quantitative YES/NO criterion was employed for selecting hits, and the hit compounds of this experimental screening were named MOA compounds (MOA1-MOA11). Of note, throughout the additional assays performed, as mentioned in the results section, the MOA compounds that showed unfavorable results were discarded as the drug discovery process progressed.

However, it is important to note that TSA analysis offered only indirect indications of ligand interaction and the identified stabilization observed did not directly correspond or rank according to their expected binding affinities. This is due to the additional combined influence of several factors, such as the concentration of free ligand, ligand binding enthalpy and binding heat capacity (185). Therefore, in order to validate the hits and assess target engagement, binding affinities were determined by ITC assays both in the presence or the absence of zinc (i.e., with both the zinc-free and zinc-bound forms). Among all the initial hit compounds, while several of them displayed insufficient binding affinities, MOA4, MOA9, and MOA10 compounds exhibited significant binding to both proBFT-3 and BFT-3 (K_d values for proBFT-3: 7.8 µM (-zinc, MOA4) and 20 µM (+zinc, MOA4), 8.5 µM (-zinc, MOA9) and 14 µM (+zinc, MOA9), 8.7 µM (-zinc, MOA10) and 22 µM (+zinc, MOA10); and K_d values for BFT-3: 8.1 µM (-zinc, MOA4) and 16 µM (+zinc, MOA4), 9.1 µM (-zinc, MOA9) and 15 µM (+zinc, MOA9), 10 µM (-zinc, MOA10) and 19 µM (+zinc, MOA10)), therefore with K_d values in the micromolar range. A higher affinity for both proBFT-3 and BFT-3 was in general observed when the zinc was absent.

After these ITC assays, a commercial kit called EnzChek protease assay kit was used to evaluate the potential *in vitro* inhibitory effect of the compounds, which allowed measuring the protease activity of the protein in the presence of MOA compounds by fluorescence. At higher compound concentrations, the emitted fluorescence was lower, so that a dose-dependent inhibition by the compounds was observed. To verify the results obtained in the assays with the substrate BODIPY-casein, several assays were proposed to evaluate the inhibitory potential of these compounds in cell-based assays. For this purpose, first, the intrinsic cytotoxicity of the compounds was determined in HT-29 cells, so that it was possible to confirm that these MOA compounds were generally not cytotoxic at the concentrations used, because most of compounds had a CC_{50} higher than 400 µM. Importantly, the compounds selected through an HTS procedure against a non-human target protein, as in the case of proBFT-3 of *B. fragilis*, should be as non-cytotoxic as possible against human cells, since the side effects subsequently produced by these compounds will be less likely to occur. In fact, in most cases it was not possible to observe differential toxic effects depending on the cell line. This is an important aspect from the point of view that they did not produce toxic effects either in colon-derived cells

(HT-29), where the real damage produced by the BFT-3 toxin occurs, and therefore where these compounds should have their effect, nor in epithelial cells (HeLa), which are cells more distant from the cell type of interest. Following this intrinsic cytotoxicity determination, other cell-based assays were performed. Firstly, it is well known that BFT-3 produces morphological changes after being added to cell cultures, such as cell rounding and dissolution of cell clusters (42). Hence, it was analyzed whether MOA compounds could reverse the morphological effect that BFT-3 toxin produced in the cells. Unfortunately, since this technique was qualitative and subjective, and the morphological changes were evaluated by optical microscopy, it was not possible to obtain reliable results confirming that these compounds decreased the levels of cell rounding and dissolution of cell clusters. Therefore, it was necessary to employ a technique that could more accurately determine whether the compounds had an inhibitory effect on the BFT-3 toxin in cell cultures. Thus, for this purpose western blot assays were carried out in HT-29 cells treated with both BFT-3 and MOA compounds. In these experiments, some MOA compounds (MOA4, MOA9, MOA10 and MOA11) exhibited the capacity of being able to reduce the levels of the E-cadherin processing at concentrations lower than their estimated CC_{50} values, maintaining most of E-cadherin in its unprocessed form. Others, such as MOA6 and MOA7 were able to inhibit E-cadherin processing only in some assays, with high variability, but these compounds were not excluded for further assays since the variability in cellular assays may be due to the metabolic/growth status of the cells in each experiment. Therefore, these compounds reduced the ability of BFT-3 to cleave E-cadherin, thus demonstrating their inhibitory efficacy in both *in vitro* cleavage assays using BODIPY-casein and cell-based assays monitoring E-cadherin processing.

The following step was to better understand which residues were involved in the ligand-protein binding. In these assays, additional evidence for target engagement and the allosteric mechanism of inhibition of these most promising compounds was obtained through X-ray crystallography. It is important to note that crystallizing the partially unfolded zinc-free state of proBFT-3 and BFT-3 would be impossible, potentially explaining the necessity for $ZnCl_2$ during crystallization. However, given that the identified compounds bound to both the zinc-bound and zinc-free states of proBFT-3 and BFT-3, the determination of the complex structures involving MOA4, MOA9 and MOA10 compounds with proBFT-3 would provide valuable information, and fortunately, it was successful. These complex structures revealed that these ligands bound to a well-defined pocket located at the upper rear of the CatD, called exosite. Notably, despite the spatial distance of the exosite from the active site, the binding of these inhibitors induced a destabilization of the active site, as evidenced by reduced zinc occupancies and increased temperature factors. Therefore, a new exosite playing an essential role in influencing geometrically the active site was discovered, and these compounds exerted the inactivation of proBFT-3 through a novel mechanism for allosteric inhibition.

Noteworthy was the fact that the identified compounds bound to both the proBFT-3 zymogen and the active BFT-3 protein, and they interacted with both the zinc-free and zinc-bound forms, although their affinity for the latter were lower as observed in the ITC assays. Interestingly, the binding energy to the zinc-bound enzyme exhibited a less favorable Gibbs energy, but this was accompanied by a more favorable enthalpic interaction, which was, however, outweighed by a less favorable (or more unfavorable) entropic term. Importantly, the competitive interaction between the active-site metal (and thus the fully folded protein) and the identified inhibitors (which favor the zinc-free partially unfolded state) was also evident in the reduced zinc occupancy in the determined crystal structures of the complexes. Of note, this competitive interplay was observed even in the presence of 5 µM $ZnCl_2$ during the crystallization process. The crystal structures also explained the similar affinities exhibited toward both the active and zymogen forms of BFT-3 due to the fact that the identified exosite was situated on the opposite side of the CatD and thus distal to the PD. Importantly, since the identified compounds were capable of binding to both the zinc-free and zinc-bound forms of BFT-3, their overall inhibitory activity likely relied on a combination of two aspects. First, there was an indirect reciprocal influence, where substrate and inhibitor reciprocally excluded each other by selectively interacting with different zinc-ligation macrostates of BFT-3, namely the zinc-bound and the zinc-free states, respectively. And second, a direct reciprocal competitive effect, in which substrate and inhibitor reciprocally excluded each other by interacting with different conformational microstates of the same macrostate, where the inhibitor distorted the catalytic process and hampered substrate binding and/or catalysis.

Therefore, MOA4, MOA9 and MOA10 showed promising results to be advanced for further assays in order to analyze their capacity as antibacterial compounds. For these reasons, the biological effect *in vitro* of these compounds was characterized in *B. fragilis* cultures. Analysis of MIC_{90} determination showed that MOA4, MOA9 and MOA10 compounds had antimicrobial effects on both, enterotoxigenic and nontoxigenic, strains of *B. fragilis*. MOA4 was the compound which reduced in a larger extent the bacterial growth both in MIC_{90} and TKA assays, therefore being the most effective compound against both strains of *B. fragilis*.

Nowadays, due to the previously mentioned emerging AMR, there are limited therapeutic options for treating multidrug-resistant organisms. In this context, synergy studies and combination therapies are commonly employed in this attempt to manage bacterial infections and improve therapeutic strategies. Therefore, it was analyzed if the most effective compound, the MOA4 compound, presented synergy effects when combined with some of the antibiotics currently used in clinic (FOX and MTZ). Analysis of synergy assays revealed that the combination of MTZ with MOA4 compound suggested a possible synergy effect between these two compounds. However, a purely synergy effect was not obtained because the combination of both compounds did not reduce ≥4-fold the MIC_{90} of MOA4 in the presence of MTZ. Therefore, the findings required further analysis in order to evaluate this possible synergistic effect. For instance, checkerboard assays (CBA) could be performed, which is one of the best known tests to determine the *in vitro* activity of a combination of antibiotics (228). Additionally, other assays could be performed, such as performing structural modifications in MOA4 for obtaining a better interaction effect between MOA4-derivates and MTZ, which could lead to a purely synergy effect. Through these studies, a new therapeutic option could be contemplated in the future, in which the current levels of MTZ used in the treatment of human infections by this pathogen could be reduced, as well as a new therapeutic framework in multidrug-resistant strains could be analyzed and considered.

To confirm the biological effect of these compounds in *B. fragilis* cultures, further steps were needed. As previously explained, the use of the *G. mellonella* waxworm as an infection model organism for studying many pathogens, bacterial load, immune response of infected larvae and searching for new drugs, has been greatly increasing in the last years (125,126). Nowadays, and very importantly, there are not published studies related to *G. mellonella* as infection animal model for studying *B. fragilis* infections. Therefore, in this Doctoral Thesis the implementation of a new infection animal model using these worms for studying infections related to this pathogen, as well as for validating previously identified compounds in experimental screenings, was one of the main important aims proposed.

For this objective, it was initially necessary to carry out an exhaustive learning process about the life cycle of these worms, taking into account the requirements to maintain them in a good health status, as well as to learn how to perform the punctures on these worms in the best possible way, i.e., avoiding any damage to the animal. In this context, an important aspect to highlight was that, in all experiments related to *G. mellonella*, two injections have been established in the experimental protocol to properly compare results between all conditions, one corresponding to the infection and the other to the treatment. Furthermore, all the worms in the same Petri dish were subjected to the same condition of the experiment and they were similar in weight and stage of life cycle. However, being a living model there were numerous factors that were difficult to control, such as transport, diet and injection-related problems, which could influence on their health status. In fact, it was of utmost importance, being one of the main aims of this Doctoral Thesis, to establish the appropriate conditions for the assays, since currently one of the main limitation in using *G. mellonella* as animal model is the absence of standardized procedures (126,133). For instance, inconsistent protocols in obtaining and storing them in the presence of antibiotics or hormones from bait stores or pet shops, lead to variable results (162). Our suppliers, i.e., Harkito Reptile, theoretically do not introduce hormones in the larvae. Additionally, results obtained in other studies indicated that the maintenance of larvae in the presence or absence of nutrition during the assays could significantly affect the development of the immune response, influencing hemolymph volume and hemocyte concentration. For instance, some studies have revealed the impact of different diets on larval survival post-infection by various pathogens such as *S. aureus*, *E. coli* and *C. albicans* (132,163). Consequently, in research studies employing *G. mellonella* larvae for analyzing the virulence of fungal or bacterial pathogens, it is of utmost importance to specify whether feeding is provided for larvae or not, as well as the kind of material supplied. These differences may affect the results obtained and may influence inter-laboratory comparisons (163). Commonly, larvae are not fed during infection experiments, and pre-infection starvation is recommended (122,141). However, recent studies have indicated that the deprivation of *G. mellonella* larvae of food throughout the experimental assays led to a decrease in immune responses and a possible increased vulnerability to infections (164). Of note, in our case, it was observed that the food supply gave a greater variability in the results obtained among treated worms of the same condition, i.e., from the same Petri dish. That was probably because the worms that were infected with *B. fragilis* or treated with MOA compounds, ate less compared to the controls (because of the remaining food left over in the Petri dishes), so that the controls showed better survival rates. Therefore, in the experiments carried out with *G. mellonella* in this Doctoral Thesis, the worms were not fed. However, because diet and immune response are linked, it could be really interesting to monitor larvae mortality with and without food intake during infection in future experiments.

After this learning and standardization period, it was first necessary to evaluate the intrinsic toxicity of the MOA4, MOA9 and MOA10 compounds on *G. mellonella* larvae. Analysis of survival rates on *G. mellonella* larvae after administering the compounds at different concentrations showed that MOA4 was the less toxic compound, and worms tolerated doses around 400 µM. In cases of MOA9 and MOA10, *G. mellonella* larvae were more susceptible for them, being the appropriate doses around 100-200 µM. Therefore, as expected, the effect was dose-dependent, so that the higher the concentration of the compounds, the lower the survival rate of the worms.

After stablishing the doses of MOA compounds required or appropriate, it was necessary to determine the number of CFU/mL needed for inducing an infection process in *G. mellonella* worms. Analysis of the survival rates confirmed that at higher concentration of CFU/mL, lower survival rates were achieved compared to the controls (*G. mellonella* larvae without a *B. fragilis* infection). Thus, in the case of the nontoxigenic *B. fragilis* strain, it was observed a reduction of survival rates for doses between 10^8-$2{\cdot}10^8$ CFU/mL at 24-48 h; and in the case of the enterotoxigenic *B. fragilis* strain at the same time, the depletion of survival rates was observed for 10^8 CFU/mL. Hence, it was clearly observed that the higher the CFU/mL of bacteria, in both strains, the lower the survival rates of the worms with respect to the control. Therefore, there was a dose-dependent correlation in the amount of CFU/mL employed, thus successfully developing a new infection animal model in *G. mellonella* larvae for *B. fragilis*. These results imply, firstly, the verification that *G. mellonella* is a good model organism to study bacterial infections, since by using these worms, it is possible to clearly evaluate the infection process produced by a bacterium. Secondly and more importantly, a new animal infection model has been developed, which represents promising novel ways to evaluate the effects produced by a pathogen, without the need to use, at a first stage, vertebrate animal models. At this point it is also important to note that an exact dose of CFU/mL necessary to produce an infection was not intended to find, but to confirm that at high concentrations of these bacteria, the infection was indeed produced in comparison with lower doses. Notably, the same dose may produce a different effect in each experiment since there were a large number of variables, previously detailed, and also other important factors. For instance, the breeding of *G. mellonella* larvae was not carried out in our laboratory, but they were administered by the commercial company Harkito Reptile. That is why, although the assays were only carried out if the larvae arrived in good health conditions, they never arrived at the same point in the developmental stages, nor in the same conditions. In addition, the temperature and transport issues were also key factors in the general health condition of the worms, which were variables that cannot be controlled but that could significantly affect the quality of the larvae. Another important factor to take into account was that, because *B. fragilis* are anaerobic bacteria, part of them were possibly affected when the inoculum was prepared for the injections. This process was not exactly the same in all the preparations; therefore, a greater or lower number of bacterial cells could have been exposed to oxygen. Of note, because these microorganisms are anaerobic represented an additional challenge when establishing the conditions for the assays, since it was of vital importance to prepare the bacterial cultures of *B. fragilis* and the injections into *G. mellonella* as quickly as possible. However, this was not a problem for the development of this infection model, mainly because although they are anaerobic bacteria, *B. fragilis* is relatively aerotolerant. This aerotolerance means that small concentrations of oxygen during short periods are not an impediment for their culture, compare to other strict anaerobes such as some species from *Fusobacterium* (229) and *Peptostreptococcus* (230) genus. In fact, as mentioned in the Introduction of Chapter I, *B. fragilis* is able to survive in the mucosa, where the typical concentrations of oxygen are higher than in the gut, which is a highly hypoxic tissue compared to other tissues, and also to induce bacteremia (41,231). Additionally, unlike other anaerobic microorganisms, *B. fragilis* can colonize the colon in the absence of facultative anaerobes, which are the microorganisms that could colonize gastrointestinal tracts of animals to reduce the partial pressure of O_2 before the subsequent strict anaerobe colonization (41). This is further supported by the fact that other anaerobic microorganisms have been previously employed in bacterial infection studies using *G. mellonella* as animal model, such as the pathogen *Clostridium perfringens* (167).

Next assays were focused on analyzing whether after infecting *G. mellonella* larvae with *B. fragilis* cultures, and treating them with MOA4, MOA9 and MOA10 compounds, a reduction in the infection process could be achieved. Ideally the control, being infected but untreated, should be the one with the lowest survival rate. In the case of MOA4 for both bacterial strains, a much higher survival rate than the control was observed, indicating that the infection was reversed when the worms were treated with MOA4. In the case of MOA9, differences were observed between strains, since less differences were observed in the nontoxigenic strain with respect to the control. However, in the enterotoxigenic strain, a higher survival rate of the worms was observed, so it seems that MOA9 could be a strain-dependent compound. In case of MOA10 compound, there was a slight increase in survival rates in the nontoxigenic strain, but less than with MOA4, and the improvement in survival rate was clearer in the enterotoxigenic strains, being also a strain-dependent compound. Due to the fact that MOA4 was less toxic in *G. mellonella* worms, and it also produced higher inhibition in bacterial cultures growth (MIC_{90} assays), it was tested at a lower concentration in worms to analyze whether similar survival rates could be observed at a lower dose. Of course, the lower the dose required for a compound, the greater the potential use that can be expected in future assays. Effectively, the 100 µM MOA4 treatment increased the survival rates in larvae infected with both strains, indicating that lower concentrations of this compound were enough to produce the same improvement in the larval survival rate. Therefore, it was verified that MOA4, MOA9 and MOA10 compounds had an antibacterial effect in *B. fragilis* infections in *G. mellonella*.

Notably, this larvae model will undoubtedly not replace mammalian trials in studies related to infection processes nor host response mechanisms. However, this *G. mellonella* model could be an additional potent tool to analyze the effect of antimicrobials and compounds against pathogens before advancing to mammal-related studies, which have higher costs and are limited by ethical regulations. Importantly, in order to have a better use of *G. mellonella*, creating stock centers for *G. mellonella* larvae similar to those for *D. melanogaster* could be highly useful for researchers. For instance, a recent development by Biosystems Technology Ltd has introduced standardized *G. mellonella* larvae called TruLarv^{TM}, which are defined by age and weight. Moreover, their breeding conditions are normalized without the use of hormones or antimicrobials, leading to more consistent results (232–234). However, a notable problem is the high cost associated with these standardized larvae, in comparison with other suppliers. Furthermore, in the last 10 years, *G. mellonella* entered the omics era, including both transcriptomic and proteomic researches, which is contributing to a better understanding of this model organism. The initial transcriptomic analysis of immune system genes led to the identification of several genes responsible for immune proteins and effector soluble molecules (148). Moreover, the genome of *G. mellonella* was sequenced (235) and a transcriptome database for these worms was established (236). These initial datasets could potentially serve as a shared-file database comprising genomic, transcriptomic and proteomic projects. Of note, further assays in our group are being focusing on analyzing variations in the transcriptome of *B. fragilis* after the treatment with MOA4, MOA9 and MOA10 compounds. The aim of these assays is to evaluate the mechanism by which these compounds inhibit the bacterial growth of *B. fragilis.* In particular, by analyzing alterations in patterns of gene expression, whether this inhibition is due solely to their effect on the BFT-3 protein, or whether there are other key molecular mechanisms involved in bacterial growth inhibition. This is an important aspect since, as already mentioned, the MOA4 compound seemed to have a similar effect between the two strains of *B. fragilis*. Consequently, such result was reflected in the data obtained in the assays with this compound in *G. mellonella*, being only slightly higher the survival rate of the worms infected with the enterotoxigenic strain in comparison with the nontoxigenic strain. Therefore, it seems that very likely the MOA4 compound not only exerts an inhibitory effect on BFT-3, but also inhibits other key processes associated with bacterial proliferation, thus increasing the survival rates of *G. mellonella* comparing to non-treated worms. Notably, MOA9 and MOA10 compounds showed some differences between strains, especially MOA9, which seemed to have a lesser effect on worms infected with the nontoxigenic strain. This could be explained by the fact that these compounds do not exclusively inhibit BFT-3, but also would be implicated in other mechanisms, and therefore the ability of both strains to proliferate is lower, providing higher survival rates than controls. Thus, the growth capacity of enterotoxigenic bacteria could be doubly affected, and although BFT-3 is a virulence factor that is not essential for the bacteria, it could facilitate the invasion process. Therefore, the involvement of MOA4, MOA9 and MOA10 compounds in additional molecular mechanisms could explain the slightly higher worm survival rates in those treated with the enterotoxigenic strain, but also the reduction observed in the nontoxigenic bacterial proliferation. Furthermore, as the survival rates of infected larvae treated with MOA4 were higher than in the other two MOA compounds, the effect that MOA4 induced in *B. fragilis* proliferation was more relevant. This fact is also confirmed by the results obtained in the MIC_{90} determination, where the MOA4 was the most effective compound in reducing the viability of both strains of *B. fragilis*.

At this point, it is really important to highlight that MOA4, i.e., flumequine, is in fact an antimicrobial that was employed to combat bacterial infections (237,238). Briefly, this compound is a first-generation synthetic fluoroquinolone primarily used in veterinary for the treatment of enteric infections affecting the intestinal tract, and also infections in the pulmonary and urinary tract (239). Its use was approved to treat, among other animals, cattle, swine, chickens, and fish, but only in a limited number of countries (240–244). In addition, in some European countries, like France, it was occasionally administered under the name Apurone for treating human urinary tract infections. However, the marketing authorization of flumequine was suspended for clinical use throughout the EU some years ago, and it is no longer commercially available. This was because the FDA considered that it may have serious side effects by affecting the musculoskeletal system (in particular the tendons), the nervous system, and even significant adverse reactions at the ocular level. Its known mechanism of action involves interfering the enzymes responsible for causing DNA to unwind and duplicate, thus killing bacteria. Specifically, flumequine is effective against both Gram-positive and Gram-negative bacteria, inhibiting DNA gyrase, a type II topoisomerase, and topoisomerase IV, enzymes necessary to separate bacterial DNA, thereby inhibiting cell division (245–247). Importantly, this mechanism has the potential to influence also the replication of mammalian cells. Therefore, the studies carried out in this Doctoral Thesis have revealed that in addition to this known mechanism of action, flumequine affects bacterial growth of *B. fragilis* cultures through a new mechanism of action, in which the binding, and subsequent inhibition of activity, to the BFT-3 toxin occurs. A comprehensive knowledge of the different molecular mechanisms that a compound exerts is of enormous importance to understand how the observed effect in a system is produced, as well as to implement future screening programs for compounds and to develop treatments for the disease under study, as well as to prevent unwanted effects. Therefore, regarding these findings, it could be very interesting to design and synthesize MOA4 derivates with the aim of trying to achieve a specific inhibition of the BFT-3 toxin. In this way, these modified compounds could not affect, or affect to a lesser extent, DNA gyrase and topoisomerase IV, in order to reduce the possible side effects in humans. On the other hand, obtaining new derivates by chemical modifications would be even more interesting for the cases of MOA9 and MOA10 compounds. This is due to the fact that, although both compounds showed antibacterial properties on both strains of *B. fragilis*, the nontoxigenic strain produced a slightly higher infection in *G. mellonella*, thus seeming clearer the direct effect that these two compounds have towards BFT-3. Thus, by obtaining more specific and potent compounds, an improvement in the treatment of infections produced by ETBF in humans could be achieved.

In addition to these transcriptomic assays, western blot assays are being performed to verify that the strains used still maintain null expression (nontoxigenic strain) and expression (enterotoxigenic strain) of the BFT-3 toxin after their purchase several years ago. As most commercial antibodies are polyclonal antibodies, it has not been possible to obtain yet a protocol in which the bacterial lysate provides clear western blot results, i.e., achieving only BFT-3 detection. Instead of that, a set of bands is being obtained in all conditions tested, indicating that the antibodies used to date have very low specificity.

In summary, BFT-3, in conjunction with the other two isoforms BFT-1 and BFT-2, is the only known virulence factor in ETBF, playing a key role in disrupting the colonic barrier and producing inflammation of the colon during infections. Additionally, BFT-3 is overexpressed in the mucosa in inflammatory bowel disease and colorectal cancer patients. Therefore, through a comprehensive approach employing an extensive compound screening, biophysical assays, *in vitro* and cell-based activity assays, and X-ray crystallography, three FDA-approved small molecule drugs: flumequine (MOA4), foliosidine (MOA9), and hesperetin (MOA10), which demonstrated the capacity to target and inhibit BFT-3 in a dose-dependent manner, were identified. Furthermore, flumequine, foliosidine and hesperetin have antibacterial activity *in vitro* in *B. fragilis* cultures, displayed low toxicity in the *G. mellonella* model and were able to reduce the *B. fragilis* infection in infected *G. mellonella* larvae. Therefore, they show great promise to be either directly repurposed for preventive or therapeutic treatment of *B. fragilis* chronic infection, diminishing the risk of intestinal inflammation and colorectal cancer development, or, given their relatively small molecular mass, to be further optimized to improve their affinity, selectivity, or bioavailability, potentially increasing their efficacy. Furthermore, a new molecular mechanism of action has been attributed to flumequine, a compound with antibacterial activity previously used to treat certain bacterial infections. Finally, this is the first report of *B. fragilis* infection in *G. mellonella* as animal model, thus successfully developing a new infection animal model in *G. mellonella* larvae for *B. fragilis* infections, and the results suggest that this new infection animal model could act as a good host to discover new antimicrobials against this pathogen.

6. CONCLUSIONS

The following **conclusions** are derived from **Chapter I - Drug discovery for inhibiting BFT-3: from a chemical screening to an animal model**:

1. The regulation of the conformation and activity of BFT-3 depends on the zinc, being this metal a key cofactor in the protein stability.

2. Eleven known FDA-approved drugs (MOA compounds), identified by a high-throughput screening procedure, target the partially unfolded state of BFT-3.

3. MOA compounds show high affinity for BFT-3 with K_d values in the micromolar range, inhibit the protease activity in a dose-dependent manner, with low cytotoxicity in eukaryotic cell lines.

4. MOA4, MOA9 and MOA10 reduce the ability of BFT-3 to cleave E-cadherin in HT-29 cells.

5. MOA4, MOA9 and MOA10 exert allosteric inhibition by binding to a well-defined pocket (exosite) located at the upper rear of the CatD of BFT-3.

6. MOA4, MOA9 and MOA10 compounds have antibacterial effects on *B. fragilis*, reducing the bacterial growth of both enterotoxigenic and nontoxigenic strains.

7. A new animal model in *G. mellonella* larvae, implemented and standardized within this Doctoral Thesis, allows to study *B. fragilis* infections and to validate previously identified compounds in experimental screenings.

8. MOA4, MOA9 and MOA10 compounds show antibacterial effects in *B. fragilis* infections in *G. mellonella*, displaying low toxicity and improving infected-larvae survival rate.

9. MOA4 (flumequine), a well-known antimicrobial, also exerts its antibacterial activity by a new molecular mechanism of action by targeting BFT-3 protein.

7. CONCLUSIONES

Las siguientes **conclusiones** se obtienen del **Capítulo I - Descubrimiento de fármacos para inhibir BFT-3: del cribado de compuestos al modelo animal**:

1. La regulación de la conformación y actividad de BFT-3 dependen del zinc, siendo este metal un cofactor clave en la estabilidad de la proteína.

2. Once fármacos aprobados por la FDA (compuestos MOA), identificados mediante un procedimiento de cribado de alto rendimiento, tienen por diana el estado parcialmente desplegado de BFT-3.

3. Los compuestos MOA muestran alta afinidad por BFT-3 con valores de K_d en el rango micromolar, inhiben la actividad de la proteasa de forma dosis-dependiente, con baja citotoxicidad en líneas celulares eucariotas.

4. MOA4, MOA9 y MOA10 reducen la capacidad de BFT-3 para escindir E-cadherina en células HT-29.

5. MOA4, MOA9 y MOA10 ejercen inhibición alostérica al unirse a una cavidad bien definida (exositio) alejada del dominio catalítico de BFT-3.

6. Los compuestos MOA4, MOA9 y MOA10 tienen efectos antibacterianos frente a *B. fragilis*, reduciendo el crecimiento bacteriano de cepas enterotoxigénicas y no enterotoxigénicas.

7. Un nuevo modelo animal de larvas de *G. mellonella*, implementado y estandarizado en esta Tesis Doctoral, permite estudiar infecciones por *B. fragilis* y validar compuestos previamente identificados en cribados experimentales.

8. Los compuestos MOA4, MOA9 y MOA10 muestran efectos antibacterianos en infecciones por *B. fragilis* en *G. mellonella*, presentando baja toxicidad y mejorando la tasa de supervivencia de las larvas infectadas.

9. MOA4 (flumequina), un antimicrobiano ya conocido, también ejerce su actividad antibacteriana mediante un nuevo mecanismo de acción molecular dirigido contra la proteína BFT-3.

CHAPTER II. Drug discovery for inhibiting NUPR1: from a chemical screening to an animal model

CONTRIBUTIONS

In Chapter II of this Doctoral Thesis, it is important to highlight the consolidated collaboration between my research group and other research groups for many years to carry out multidisciplinary studies in order to obtain results of high scientific quality. In Chapter II, the assays performed by the PhD student will be detailed, but it is important to keep in mind that these assays are part of more complex and ambitious studies.

The following is a list of the principal researchers involved in this collaboration:

- Juan L. Iovanna, Patricia Santofimia-Castaño and Xi Liu (*Centre de Recherche en Cancérologie de Marseille (CRCM), INSERM U1068, CNRS UMR 7258, Aix-Marseille Université and Institut Paoli-Calmettes, Marseille, France*).
- José L. Neira (*Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Elche, Alicante, Spain*).
- Bruno Rizzuti (*CNR-NANOTEC, Licryl-UOS Cosenza and CEMIF.Cal, Department of Physics, University of Calabria, Cosenza, Italy*).
- Ling Peng (*Aix-Marseille Université, CNRS, Centre Interdisciplinaire de Nanoscience de Marseille, UMR 7325, Marseille, France*).
- Yi Xia, Zhengwei Zhou and Li Zexian (*Chongqing Key Laboratory of Natural Product Synthesis and Drug Research, School of Pharmaceutical Sciences, Chongqing University, Chongqing, China*).

1. INTRODUCTION

1.1. Pancreatic cancer

Among gastrointestinal diseases, pancreatic cancer (PC) is one of the pathologies of greatest concern to the scientific community and the general public, as it is one of the most aggressive and with the worst prognosis, given the poor response to current treatments, the high probability of recurrence and late detection (11). Despite ranking 12th in the classification of the most frequently diagnosed tumors, although PC is not among the cancers with the highest incidence, its enormous and rapid aggressiveness and development make the mortality rate disproportionately high (7). More specifically, according to data from the International Agency for Research on Cancer, PC was estimated to cause 466,003 deaths in 2020, accounting for 4.7% of cancer deaths worldwide, and being the 7th leading cause of death in the world. Pancreatic ductal adenocarcinoma (PDAC) is the most common type of PC, representing 90% of the cases. The incidence of PDAC is estimated to be increasing by 0.5-1% per year, becoming the 2nd leading cause of cancer-related death by 2030 (**Figure 61**). According to the Spanish Network of Cancer Registries (REDECAN), 9,252 new cases of PC were detected in Spain in 2022 (248). Other types of pancreatic neoplasms include acinar carcinoma, pancreatoblastoma and neuroendocrine tumors (249–251).

The 5-year survival rate has increased from 5.26% in 2000 to 10% in 2021 thanks to multi-agent therapies, despite the fact that there are currently no effective diagnostic screenings and that most PC patients present with non-specific symptoms at advanced stages of the disease, where curative surgery is no longer possible (252–254). When detected, 50% of patients have metastatic disease, 10-15% of them have localized disease that is amenable to surgery, and 30-35% of affected people have locally advanced disease that mostly cannot be surgically removed due to the high degree to which the tumor-vascular is affected (249). The huge problem with the late diagnosis lies in the fact that the rate of metastasis in PDAC is quite high: the liver (90%), lymph nodes (25%), lungs (25%), peritoneum (20%) and bones (10-15%) being common sites of metastasis in this type of cancer (252).

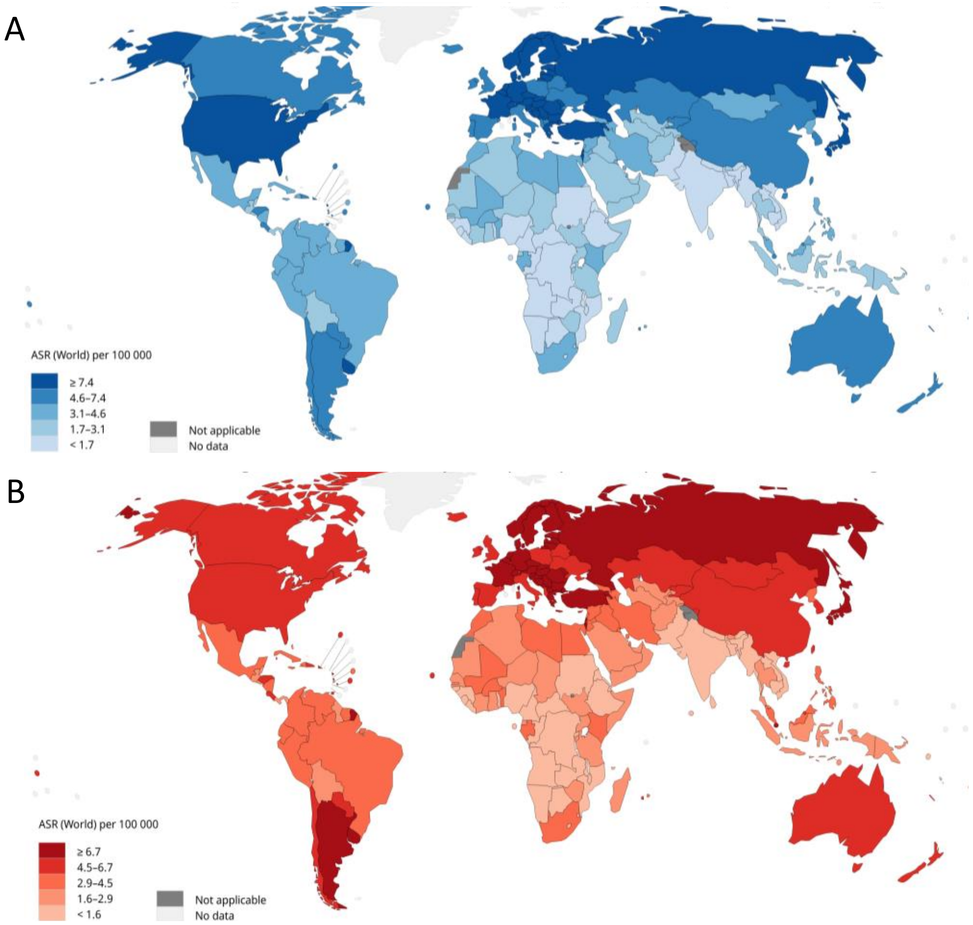


Figure 61. Age-standardized PC (A) incidence and (B) mortality rates for countries in five continents according to the latest WHO International Agency for Research on Cancer figures. Figure from: Globocan 2020. Graph production: Global Cancer Observatory. Institutional Agency for Research on Cancer (World Health Organization).

In terms of risk factors, tobacco is the lifestyle factor with the strongest association with PDAC, followed by an excessive alcohol intake (255,256). Besides, it has been observed that chronic pancreatitis is associated with a 13-fold increased risk of developing PDAC, and obesity has also been associated with an increased risk of PDAC. Given the increased consumption of processed food, which contains more pro-inflammatory components, the incidence of PDAC in people under 30 years of age is also increasing (257,258). On the other hand, 3.8-9.7% of PDAC patients with germline genetic variants in their germline make them more susceptible to PDAC. In most cases, these variants occur in DNA damage repair-associated genes. Among the most common variants are *Breast Cancer Type 2* (*BRCA2*), *Breast Cancer Type 1* (*BRCA1*) and *Ataxia-telangiectasia mutated* (*ATM*) genes (259–261). These factors contributing to the rising incidence of PC are associated with a stressful environment, leading for instance to an increase in pro-inflammatory cytokine release or oxidative stress with ROS overproduction and fatty-acid peroxidation (262). These processes can cause damage on proteins, lipids and DNA, leading to the generation of toxic and mutagenic metabolites that disrupt cellular homeostasis (263), inducing to DNA aberrations and facilitating tumor cells to acquiring invasive and aggressive phenotypes. Therefore, these factors are involved in the instability and high levels of stress present in PDAC (264).

A complex set of multistep genetic alterations is involved in the pathophysiology of PDAC (**Figure 62**). There is a precancerous state, in which there are pancreatic intraepithelial neoplasms (PanINs), that are precancerous lesions, of which a small fraction may progress to a high degree of dysplasia and PDAC (265). In this previously mentioned precancerous state, PanINs acquire a series of cumulative genetic defects, resulting in instigating oncogenes that are responsible for both the initiation and maintenance of PDAC. Some of these oncogenes are *Kirsten rat sarcoma virus* (*KRAS*), *cyclin-dependent kinase inhibitor 2A* (*CDKN2A*), *tumor protein P53* (*TP53*), *SMAD family member 4* (*SMAD4*) and *BRCA2* (266,267).

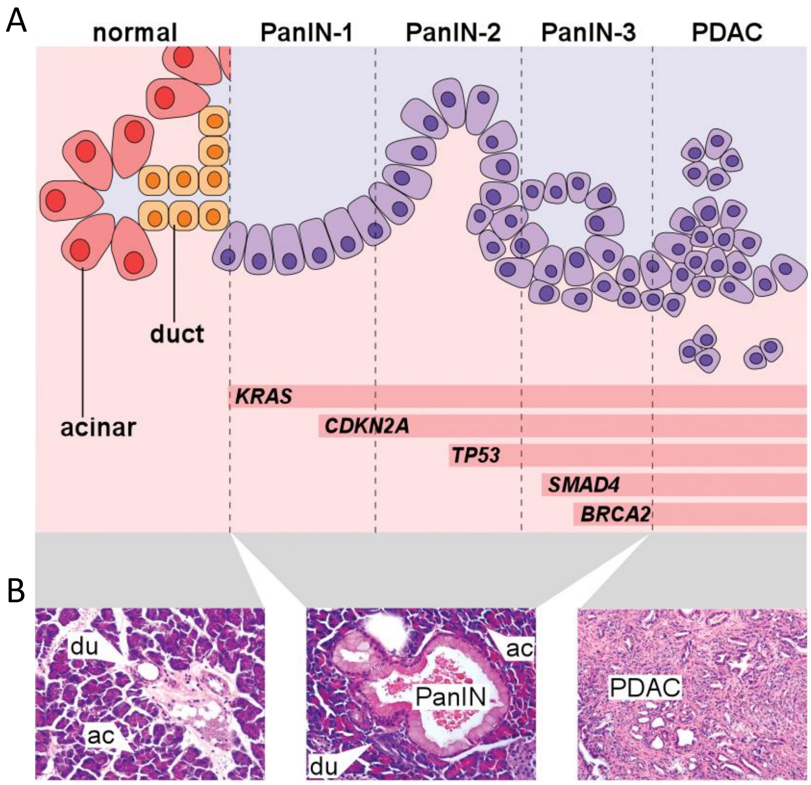


Figure 62. Representation of the pancreatic cancer progression in human and mouse. (A) Normal exocrine pancreas schematic representation cells, indicating from left to right the increasingly dysplastic PanIN formation and PDAC development. PanIN-PDAC progression is related to an increasing accumulation of genetic lesions, involving the activation of *KRAS* and loss of tumor suppressor genes including *CDKN2A*, *TP53*, *SMAD4* and *BRCA2*. (B) H&E (hematoxylin and eosin)-stained sections of mouse pancreas of the following genotypes: left (WT), middle (*Kras*^{LSL}; *Pdx1-Cre*), right (*Kras*^{LSL}; *Pdx1-Cre; p53*$^{+/-}$), illustrating normal acinar (ac) and duct (du) cells as well as an early-stage PanIN lesion and advanced PDAC. Figure adapted from (267).

As for the symptoms of PDAC, as previously mentioned they are often nonspecific, which results in a late diagnosis in most cases. Among them are decreased appetite, indigestion and changes in bowel habits. In most cases (70%) PDAC tumors begin in the head of the pancreas with often biliary obstruction, thereby producing dark urine, jaundice, weight loss, fatigue, loss of appetite and exocrine pancreatic insufficiency (268).

Regarding the treatment, PDAC is normally treated with surgery, chemotherapy, and radiation therapy, but it is rarely cured. As previously mentioned, less than 20% of PDAC patients can afford surgical removal of the tumors because detection usually occurs at an advanced or metastatic stage (269,270). This makes chemotherapy the main treatment for PDAC. The main chemotherapeutic agents are those based on DNA-damaging agents, which directly affect DNA synthesis and repair (e.g., oxaliplatin, irinotecan), and antimetabolites such as gemcitabine and fluorouracil are also employed. If metastasis is present, e.g. PDAC is at an advanced stage, multiagent cytotoxic regimens are able to improve survival rates (252,253). The current first-line chemotherapy options include gemcitabine alone or combined with albumin-bound paclitaxel (nab-Paclitaxel), or modified FOLFIRINOX (which is a combination of 5-fluorouracil, oxaliplatin, irinotecan, and leucovorin) (252,271–273).

Since there is currently no fully effective treatment for PDAC, the search for new and more effective therapeutic targets and drugs is actively progressing. As mentioned in the General Introduction, drug discovery in many cases is based on the search for compounds that are effective against proteins involved in some pathology. In this context, the development of drugs against proteins involved in cancer, whether overexpressed, underexpressed, or mutated, is and appropriate strategy to combat this deadly disease that affects a large number of people around the world. In addition, the interest in the search for drugs active against the so-called intrinsically disordered proteins (IDPs), which are biologically active proteins with undefined structure and considered undruggable until recent years, has grown exponentially, due to the fact that many IDPs are associated with key cellular processes: signaling, transcription, translation, homeostasis and cell cycle (274).

1.2. Intrinsically disordered proteins

During the 20th century, the protein structure-function paradigm was developed and consolidated, in which it was assumed that a sequence of amino acids gave rise to a 3D structure that was capable of fulfilling one or more functions. A key prerequisite of this paradigm was that, for a protein to have a function, a well-defined 3D structure was compulsory, so that the native structure of a protein was equivalent to having an ordered 3D structure. Later, examples were discovered in which unstructured segments of proteins played key roles in certain protein functions (275). In some cases, lack of electron density in certain parts of proteins could be observed, indicating that these atoms were disordered. However, there were numerous reports in which these disordered regions were crucial for function (276–278). Some of these studies made the 3D NMR technique essential for the structural determination of disordered regions of functional proteins (279).

Thanks to these advances, the intrinsically disordered proteins (IDPs) were defined, in which there is an absence of secondary structure and/or stable tertiary structure under physiological conditions. These proteins are highly abundant in nature, and they are necessary to complement the functions of ordered proteins. As mentioned above, IDPs are involved in cell signaling, regulation and control processes, as they bind to multiple biological partners, and they often do so with high specificity but very often with moderate to low affinity interactions, thus playing crucial roles in biological systems. The degree of disorder of these proteins allows them to participate in both one-to-many and many-to-one signaling events. Despite the lack of secondary and tertiary structures and the high flexibility of IDPs, they have crucial biological functions, such as molecule recognition, molecule assembly, protein modification and entropic chain activities (279–281).

Protein dysfunction can result in various pathologies as well as disorders, as proteins are highly involved in the maintenance and proper functioning of life. The fact that IDPs are in great abundance implicated in various diseases is above all due to their ability to fold, partially fold, or adopt structures at the local level upon interaction with the corresponding biological partner, their ability to have precise functional control over alternative splicing and PTMs, and the fact that IDPs are highly promiscuous, in terms of binding and their great plasticity. The acquisition of local order in their structures can be achieved by PTMs, among other factors, and these modifications can expand their biological functions (282,283).

In this context, the vast majority of the successfully targeted IDPs play crucial roles in cellular stress responses. Cancer cells, in particular, grow in a hypoxic microenvironment characterized by a limited nutrient supply, elevated mechanic strain, and various other challenging conditions, leading to the activation of stress proteins. The presence of stress proteins can help cancer cells in adapting to these harsh environmental conditions. Consequently, cancer cells that develop and grow under stressful conditions highly rely on the functionality of stress IDPs (284,285). Some examples of IDPs implicated in human diseases, associated to high levels of stress, are α-

Fetoprotein (development of cancer or fetal abnormalities), p53 (cancer development), BRCA-1 (breast cancer), α-Synuclein (synucleinopathies), Aβ and Tau protein (Alzheimer’s disease), Hirudin and Trombin (cardiovascular disease), Amylin (type II diabetes), and NUPR1 (PC) among others (280).

1.2.1. NUPR1: an overview

Nuclear protein 1 (NUPR1), also known as Com-1 (candidate of metastasis 1) or p8 (molecular mass 8 kDa) is a protein which is evolutionarily conserved among organisms such as *Drosophila, Xenopus*, nematodes and mammals, but not in yeast. The *nupr1* gene is located on chromosome 16, position p11.2 (286), and is a stress-induced gene firstly identified in the pancreas during the acute phase of pancreatitis, when the protein is activated (287). NUPR1 is a stress protein which is primarily found in the nucleus of various cancer cells, therefore playing important roles in cell stress, senescence, cell cycle regulation, cell migration and invasion, stress-related apoptosis, and metastases, among others (284). Of note, in cancer cells with low density growth, NUPR1 primarily localizes to the nucleus, while in high-density grown cells, it is distributed throughout the entire cell (288). In the recent years it has been observed that NUPR1 is overexpressed under stress conditions caused by many stimuli in a large variety of different kind of human cancers, and in most cell types, and it has been described as an important player in roles related to the development and progression of several types of cancer, such as PC (289,290), and other pathologies, including pancreatitis, inflammatory and neurological conditions, and diabetes (291), therefore having a crucial role of this protein in cancer development (292,293).

1.2.1.1. NUPR1: sequence, isoforms and structure

NUPR1 (UniProtKB O60356) is an 82-residue-long (8.8 kDa), monomeric, basic IDP without a stable secondary and tertiary structure, as it happens in IDPs (284). Until now, all biochemical and molecular biological research related to NUPR1 has exclusively focused on the shorter isoform, isoform β. However, alternative splicing of the *nupr1* gene generates an additional isoform (isoform α) which contains an extra 18 amino acids. The tissue expression pattern and potential functional differences associated with this additional isoform remain unidentified, highlighting the need for further research to explore its significance. Therefore, all the data of NUPR1 characterization concern the shorter isoform β, and the biological implications and significance of isoform α, including its expression in different tissues and any functional divergences it may possess, have yet to be explored (284,294). The canonical sequence (isoform β) is indicated in (**Table S6, Annex III**).

Regarding its structure, from the amino acids 1 to 82, the protein is considered mainly disordered (a representation of the structure prediction can be observed in **Figure 63**). Specifically, NUPR1 is a polypeptide which contains a canonical bipartite domain, which serves as a nuclear localization signal (NLS). This NLS is positively charged and spans amino acids 65 to 82, and a N-terminal PEST (Pro/Glu/Ser/Thr-rich) region, which is related to the regulation of NUPR1 protein levels by the ubiquitin/proteasome system. At its C-terminus, NUPR1 contains a basic helix-loop-helix motif (294,295). Furthermore, NUPR1 contains two critical "hot spot" regions, which have been characterized through *in silico* methods and protein engineering assays, involved in the binding to its natural partners, including, for example, DNA (without an apparent preference for a specific DNA sequence), prothymosin α (ProTα), male-specific lethal protein (MSL), and the C-terminal domain of RING1B (C-RING1B) (296–298). The first "hot spot" region is located around residues Ala33, and the second "hot spot" is located around Thr68. Both of these regions are notably hydrophobic polypeptide patches within the protein chain (299).

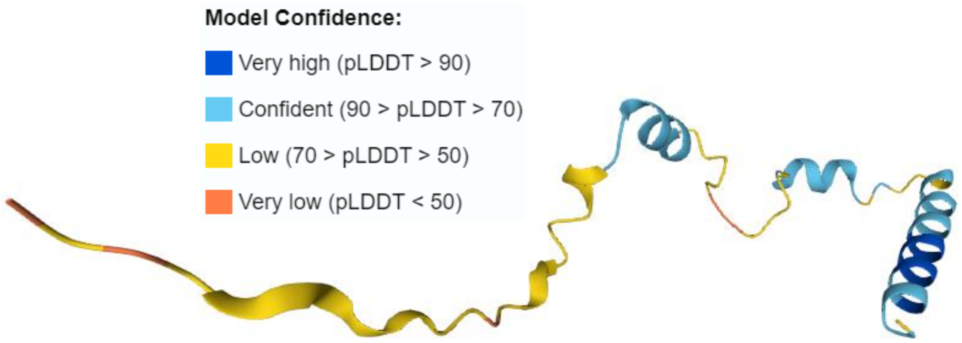


Figure 63. NUPR1 structure predicted by Alphafold tool. AlphaFold is an artificial intelligence (AI) system developed by DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) that predicts the 3D structure of a protein from its amino acid sequence. It regularly achieves accuracy competitive with experiment. For further information, see AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/). AlphaFold produces a per-residue confidence score (pLDDT) between 0-100. Some of the regions that have low pLDDT may be unstructured in isolation.

1.2.1.2. NUPR1: functions and implications

Induced NUPR1 is associated with a multitude of functions, often contradictory and challenging to comprehend. When considering collective reports, there is evidence that NUPR1 is a complex molecule, exhibiting diverse physiological and biological functions. As previously mentioned, among other important roles, NUPR1 expression is related to stress response derived from several different stimuli (such as hypoxia, apoptosis inducers, glucose starvation and several anticancer agents), interacts with DNA and controls the expression of target genes, participating in processes associated with PC: cell cycle regulation, apoptosis, metastasis and DNA repair, therefore making tumor cells genetically unstable and protecting them from apoptosis (284,294,299). Its inactivation by inhibitors or genetic mechanisms induces tumor growth arrest (292,293). Some of the key functions of NUPR1 in cells and the associated genetic interactions are explained below and a schematic outline summarizing them is illustrated in **Figure 64**.

Transcriptional regulation of NUPR1:

When there is a stress situation, that could be induced by several genotoxic signals as well as genotoxic agents, NUPR1 is typically expressed in response to them. One of the most important regulators of NUPR1 transcription is the transforming growth factor β (TGFβ). When TGFβ binds to its receptor, a canonical cascade starts, in which a heteromeric complex of phosphorylated small mothers against decapentaplegic (SMAD)-2/3 proteins and the cofactor SMAD-4 is formed, and then translocated into the nucleus. In this cellular location, it binds to the promotor at the 5'-untranslated region (5'-UTR), quickly elevating the transcription of NUPR1 (300).

DNA damage and repair:

NUPR1 protein is also related to DNA damage due to the fact that this protein has an important role influencing cancer cell resistance to metabolic stress-induced nutrient starvation, such as glucose and hypoxia. This regulation is mediated through the downstream regulation of Aurora kinase A (AURKA) expression, since its inhibition triggers a cytotoxin that can lead to DNA damage (301). In γ-irradiation-induced damage and repair, NUPR1 is also important because it negatively controls DNA repair, produced by this type of irradiation, in the presence of MSL complex subunit 1 (MSL1). This reparation is carried out by regulating histone acetyltransferase (HAT) activity. In this process, NUPR1 is also involved in the intercommunication with p53 binding protein (P53BP1) (302). Other studies indicated that inhibiting NUPR1 by compounds, such as the organic synthetic molecule ZZW-115, sensitizes cells to DNA damage, producing a reduction in the SUMOylation, a PTM involved in various cellular processes, of several proteins which are involved in DNA damage response and by inhibiting the nuclear translocation of NUPR1. This inhibition decreases the SUMOylation-dependent functions of several proteins involved in DNA damage response (292).

Cell stress and cell death:

Depending on the type of cellular stress and cell situation, NUPR1 is able to regulate cellular damage and death in different ways. It is known that this protein is involved in D9-tetrahydrocannabinol (THC)-induced cancer cell death through the downstream targeting of death inducible telomere repeat-binding factor 3 (TRB3) protein, activating transcription factor 4 (ATF-4) and the protein C/EBP homologous protein (CHOP) following ERS elevated from the synthesis of ceramide and collectively induces apoptosis (303). Besides, in hepatocellular carcinoma (HCC) cells, the NUPR1 downregulation can increase cell sensitivity to some antineoplastic treatments, for example to sorafenib, which also controls through the RELB/IER3 pathway the cell growth (304). By knockdown of NUPR1 expression, cell death can be observed, which could be reversed by incubation with necrostatin-1, but not inhibiting the caspase activity. The inactivation of NUPR1 leads to ERS in PC cells, inducing in this way a mitochondrial malfunction related to a deficient ATP production, producing as consequence cell death mediated by a programmed necrosis pathway as a consequence (305).

Cell growth, autophagy and death:

NUPR1 can influence cell cycle progression, thus promoting cancer cell proliferation. Specifically, cells can enter the S phase bypassing the G0/G1 checkpoint thanks to NUPR1 help. This escaping from the G0/G1 phase is produced due to an association that exists between NUPR1 and cyclin inhibitory proteins, producing a downregulation of p21 and p57 (306). Besides, NUPR1 interferes with FoxO3 to regulate autophagy, promoting the Bnip3 transcription in the autophagy control, which is related to a stress-dependent self-defense mechanism that facilitates cells to eliminate the toxic microenvironment (307). It has been also demonstrated that NUPR1 and ProTα, which is a protein involved in cell proliferation, chromatin remodeling, carcinogenesis, and cellular and viral transcription, form a heterodimeric complex involved in the regulation of an apoptotic cascade, the staurosporine-induced apoptosis (308).

Cell senescence:

It has been also described in pancreatic mouse cells that the KRAS-induced cellular senescence is facilitated through downregulating Dnmt1 (DNA (cytosine-5)-methyltransferase 1) expression when NUPR1 is inhibited. Dnmt1 is an enzyme that transfers methyl groups onto the DNA, increasing thus the DNA methylation, that is essential for transformation and transcription regulation, and helping the induction of KRAS-dependent PC. The way in which NUPR1 regulates Dnmt1 expression is by directly binding to the promoter and thus interfering with the transcription (309,310). Thereby, NUPR1 is an important co-operator factor with the *KRAS* oncogene, and it is already known that both factors promote PanINs both *in vivo* and *in vitro* systems (289).

Endothelial cells:

NUPR1 plays an important role in the process of endothelial apoptosis. In response to ERS, NUPR1 increases the production of the CHOP transcription factor, and through the mediation of p53, which is considered "the guardian of the genome" due to its role in conserving the genome stability by preventing genome mutation, it couples to the PUMA promoter. Then, the anti-apoptotic B-cell lymphoma 2 protein (Bcl-2) is dissociated and the pro-apoptotic BAX is upregulated, both by the action of CHOP, and also the membrane potential of the mitochondria is altered. The apoptogenic factor Cyt*c* is imported from the mitochondria to the cytosol due to the increased BAX/Bcl-2 ratio, and this factor leads then endothelial caspase-mediated cell death (311). Cyt*c* is a key player in cell fate decisions because it is implicated in many processes, such as interacting with a variety of well-known histone chaperones, related to the apoptosis process (312,313).

Influence on metabolism:

It has been described that PC-associated cell apoptosis and programmed necrosis are induced after NUPR1 depletion, which also cooperates with ERS for these actions. If NUPR1 is inhibited, the ATP production is decreased, producing a deficiency in oxygen availability (305). Furthermore, when there is a NUPR1 knock-down together with ERS, a disruption in the mitochondrial membrane occurs, followed by a calcium uptake from the cytoplasm and then leading to an alteration in the membrane permeability. This alteration results in the discharge of Cyt*c*, triggering cell death (311). On the other hand, some studies indicate that there is an association between the lack of NUPR1 and bone metabolism, which is mediated by the receptor activator of NF-κB ligand (RANKL) and sclerostin, enhancing the proliferation of osteoblasts and the downregulation of osteoclasts (314). Besides, NUPR1 interacts with p300 and Pax2, influencing in this way in the promoter activities of the glucagon gene (315). NUPR1 also induces beta cells growth in the pancreas by mediating glucose signaling or metabolism (316). Therefore, NUPR1 is an important regulator in glucose homeostasis and protein metabolism (317).

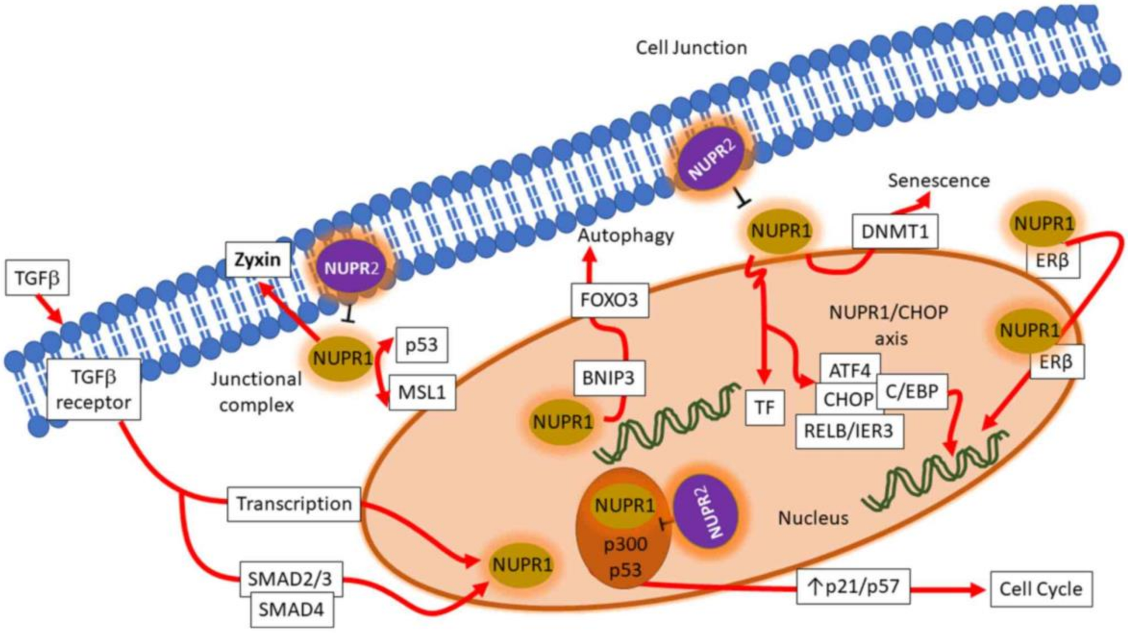

Cell Junction
NUPR2
Senescence
NUPR1
DNMT1
NUPR1
ERβ
Autophagy
Zyxin
NUPR2
TGFβ
FOXO3
NUPR1/CHOP
axis
NUPR1
ERβ
NUPR1
p53
TGFβ
receptor
Junctional
complex
MSL1
BNIP3
ATF4
CHOP
C/EBP
TF
NUPR1
RELB/IER3
Transcription
NUPR1
NUPR2
p300
p53
Nucleus
NUPR1
SMAD2/3
SMAD4
↑p21/p57
Cell Cycle

Figure 64. The key functions and interactions of NUPR1 in cells are detailed. Of note, further details of nuclear protein 2 (NUPR2) are detailed below. Figure from (291).

1.2.1.3. NUPR1-related sequences: NUPR1-like

Recent studies in the human genome have revealed the existence of a NUPR1-related sequence, consequently called nuclear protein 1-like (NUPR1L) and also known as NUPR2 (318). NUPR1L is a paralogue of NUPR1 and, as well as NUPR1, it is an IDP, but with more propensity to self-associate than NUPR1, even having some regions with conformations such as turn or helix-like motifs (319). NUPR1L has a length of 97 residues and it is involved in DNA repair and chromatin remodeling; therefore, NUPR1L also must be translocated to the cell nucleus to carry out its functions and bind to DNA. Besides, this kind of paralogue of NUPR1 is a new direct p53 target gene, meaning that *nupr1L* gene is regulated and activated after p53 activation in response to DNA damage or cell cycle arrest. It has been demonstrated that NUPR1L down-regulates at the transcriptional level the tumorigenic *nupr1* gene by the repression of the activity of its promoter. Furthermore, the forced expression of NUPR1 produces an increase in cancer cells viability, which is reduced by NUPR1L expression, meaning that there is a functional interaction between both proteins (320). In the same way than NUPR1, NUPR1L also interacts with ProTα and C-RING1B in the low micromolar range, and importantly it was also demonstrated that NUPR1L interacts with NUPR1 through the same "hot spots" of NUPR1. Therefore, the modulation of NUPR1 by NUPR1L not only takes place at the transcriptional level, but there is also direct regulation at the protein level (319). The sequence of NURP1L is indicated in (**Table S7, Annex III**).

1.2.1.4. NUPR1 and its interaction with biological partners

As discussed in this Doctoral Thesis, numerous studies have confirmed that NUPR1 interacts with a large number of biological partners. To gain a better understanding of these interactions, recent studies conducted by our collaborators determined the interactome of NUPR1 (292). Specifically, using a Flag-tagged NUPR1 fusion protein in an immunoprecipitation assay followed by liquid-chromatography tandem mass spectrometry proteomic analysis, we identified 656 proteins capable of interacting with NUPR1. Among these proteins, more than 30 were found to be components of the nuclear pore, including some importins. It is worth noting that, in addition to the importance of the NLS sequence of NUPR1, NUPR1 employs two "hot spots" around residues Ala33 and Thr68 in interactions with several partners and synthetic molecules (180,297,298).

Having a thorough comprehension of the biological partners of NUPR1, as well as of the residues involved in specific binding to them, is crucial for a better understanding of the mechanisms of action carried out by NUPR1 through its different functions, and for the development of drugs capable of inhibiting such interactions, thereby blocking the oncogenic potential of NUPR1.

1.3. Molecular mechanism of the classical nuclear protein transport and the role of Importin α

In eukaryotic cells there are multiple carrier proteins involved in the protein transport from the cytoplasm to the nucleus by recognizing specific signals on cargo molecules. This molecular trafficking is a crucial cellular process and necessary for cell survival and development. Among all carrier proteins, importins, together with nucleoporins and GTPase Ran, are the most important ones (321,322). Importin α (Impα) is one of the more relevant and presents several human isoforms. Besides, one of the necessary requirements for a protein to be transported to the nucleus is to contain a NLS within its sequence, due to the fact that the classical nuclear import pathway starts when Impα recognizes the NLS in the cargo. Most of the NLS identified in proteins comprise 8-10 sequential amino acids enriched in basic residues such as lysine and arginine, allowing specific interactions with the key residues of the inner surface of Impα (321,323). Briefly, the complex cargo-Impα-Impβ is formed (importin β: Impβ), and this ternary complex goes to the nucleus through the nuclear pore complex (NPC). Then, in the nucleus there is a protein called GTPase Ran, which is responsible for dissociating the ternary complex after interacting with Impβ. Subsequently, both Impα and Impβ are recycled back to the cytoplasm with the help of CAS exportin and GTPase Ran (**Figure 65**) (282,324).

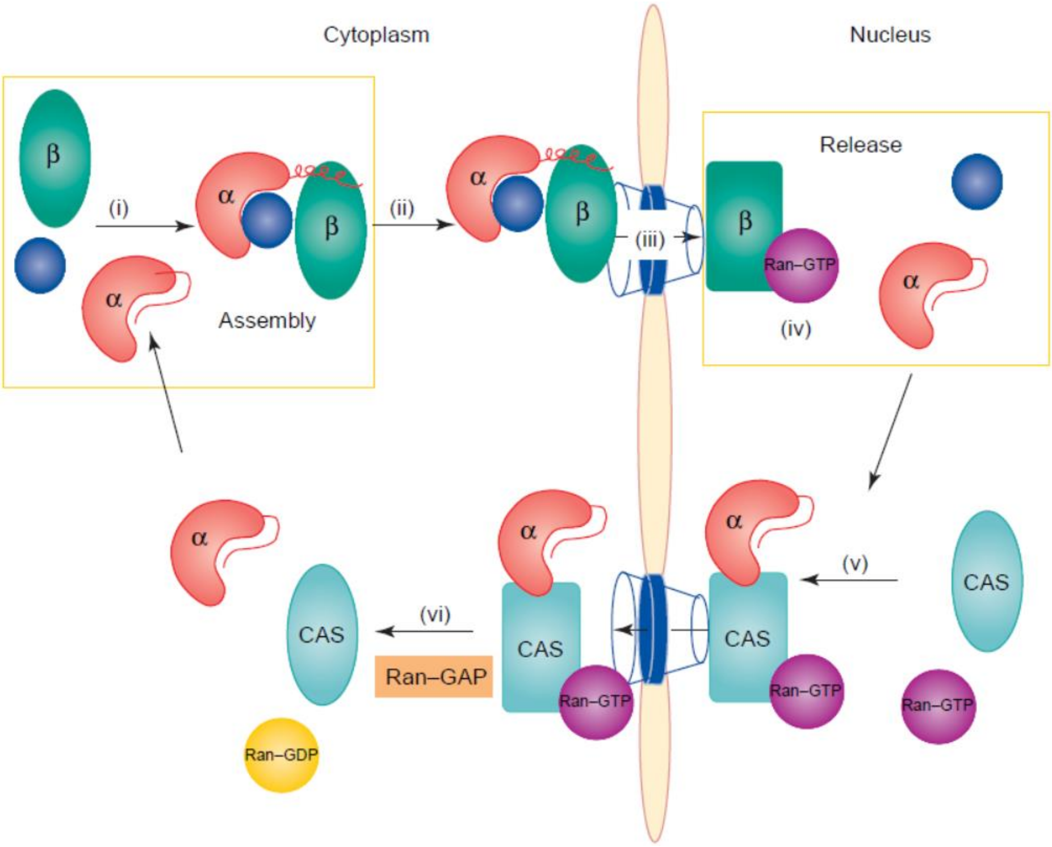


Figure 65. The nucleocytoplasmic shuttling cycle of Impα involves the following steps: (i) Impα forms a complex with Impβ and cargo (represented by blue circles), creating a ternary complex. (ii) The ternary complex docks at the NPC. (iii) Subsequently, the complex translocates into the nucleus. (iv) The binding of Ran-GTP triggers the dissociation of the ternary complex. (v) Impα then binds to the exportin CAS-Ran-GTP complex and is exported back to the cytoplasm. (vi) The dissociation of the exportin complex is initiated by the Ran-GAP-stimulated hydrolysis of GTP by Ran. This process releases free Impα into the cytoplasm, ready for another round of transport. Of note, the recycling of Impβ to the cytoplasm, as well as the transportation of Ran-GDP to the nucleus and its subsequent conversion to Ran-GTP, are not depicted in this representation. Figure from (324).

Impα was firstly identified as a crucial adaptor protein responsible for linking NLS-containing proteins to Impβ (324,325). The human genome encodes seven Impα subtypes, which are classified into three subfamilies based on their sequence similarity (323):

- α1 subfamily: importin α5, importin α6 and importin α7 (Impα5, Impα6 and Impα7, respectively).
- α2 subfamily: importin α1 and importin α8 (Impα1 and Impα8, respectively).
- α3 subfamily: importin α3 and importin α4 (Impα3 and Impα4, respectively).

All these isoforms have similar chain lengths, between 516-539 residues, and have as many as 31% identical and 50% conserved residues in their armadillo (ARM) cores (326). These different subtypes exhibit a cargo-specific affinity, allowing them to transport distinct subtype-specific molecules into the nucleus. Besides, these subtypes are expressed in a specific pattern depending on certain tissues, cell types or developmental stages, therefore suggesting that Impα subtypes play crucial roles in cell-type specific functions, cell specification and cell differentiation processes (323). On the other hand, importins have a high flexibility for the target recognition. In particular, among all the importin isoforms, Impα3 is more flexible than the others because of the high flexibility of its ARM core, and because it binds weakly to classical NLS sequences (327). The sequence of Impα3 is indicated in **Table S8, Annex III**.

Impα is a ~60 kDa protein containing three key structural domains: a N-terminal importin β-binding (IBB) domain, ARM repeats acting as internal cargo NLS-binding sites, and a C-terminal region binding to CAS, the nuclear export factor of Impα (**Figure 66**). The central segment of Impα consists of 10 repetitive motifs, each encompassing 42-43 amino acids approximately, with a relatively hydrophobic character (ARM repeats). A NLS-containing cargo molecule may bind to two different sites within the ARM repeats, which are referred to as major (ARM repeats 2-4) and minor (ARM repeats 6-8) NLS binding sites (323,328). Of note, the IBB domain provides a dual role: this domain is used for binding to Impβ to transport the complex with the cargoes through the NPC, but it is also known that the IBB domain has an autoinhibitory function for NLS-binding, because it contains an autoinhibitory sequence with a cluster of basic residues mimicking a NLS, therefore regulating the binding of NLS cargoes to the ARM domain of Impα. When Impα is not bound to Impβ, this autoinhibitory sequence within the IBB domain self-interacts with the NLS-binding pocket. However, NLS cargos can still bind to Impα due to the fact that this interaction is not extremely strong, but notably reducing their binding affinities compared to when Impβ is present (324).

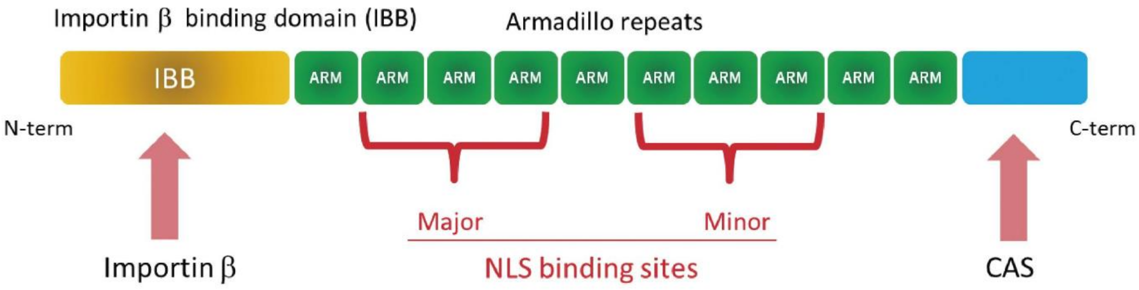


Figure 66. Scheme of the structure of Impα with the different domains. Figure from (323).

1.4. Targeting NUPR1 to treat PDAC: an overview of our research in recent years

As previously mentioned, designing drugs that specifically target IDPs poses significant challenges due to their dynamic structural nature, making traditional structure-based drug design methods impractical (274). Therefore, developing inhibitors for NUPR1 has proven to be particularly challenging due to the unique structural and functional characteristics of this protein.

For years, in our research group, in collaboration with the different groups mentioned at the beginning of Chapter II, we have carried out numerous studies to understand the role of NUPR1 in a more comprehensive way, as well as to develop drugs against this protein, applying basic research based on biophysical, biochemical, bioinformatics and biological techniques, as well as the subsequent evaluation of these compounds in more complex biological systems such as cell cultures and animal models of PC. For this reason, our studies in this area have brought to light the complexities associated with targeting NUPR1 and designing effective protein-protein interaction inhibitors, in order to block the action of NUPR1 and thus reduce the risk of PC.

At this point it is important to emphasize that all the efforts and advances that will be detailed below are undoubtedly worthwhile because: firstly, PC is a cancer with a high level of stress of different origins; secondly, NUPR1 is a stress protein and a hub protein involved in several key signaling pathways; and thirdly, NUPR1 is specifically overexpressed in PDAC relative to its null expression in normal tissues, making NUPR1 a potential therapeutic target for the development of new anti-cancer therapies in PDAC.

We have focused on adopting a bottom-up approach in designing drugs against IDPs, concretely utilizing NUPR1 as a proof-of-concept (**Figure 67**). This approach involved a multi-disciplinary effort, integrating *in silico*, *in vitro*, *in cellulo* and *in vivo* techniques that were readily available. Unlike other approaches for designing drugs targeting IDPs, our method did not rely on a complete experimental determination of the conformational ensemble of the IDP. Instead, we focused on the specific characteristics of screened compounds. Then, these compounds underwent chemical modifications and were studied through complementary methods, both *in vitro* and *in vivo*, to gain a comprehensive understanding of their mode of action. Through *in silico* assays, we employed computational tools to analyze the interactions between compounds and NUPR1. These simulations provided valuable insights into the binding mechanisms and potential binding sites. In the *in vitro* techniques, we performed experiments to characterize the mechanism of actions of compounds and their effects on NUPR1. This involved different assays and measurements to assess their binding affinities, thermodynamic properties and structural changes induced in NUPR1. Additionally, we employed *in cellulo* techniques, utilizing cell-based models, to examine the effect of compounds on cellular processes involving NUPR1. This allowed us to observe their effects within a cellular context, studying its behavior and potential downstream effects. Furthermore, in the *in vivo* assays, we evaluated the efficacy and safety of compounds using animal models. By administering compounds and monitoring their effects *in vivo*, we gained insights into their overall pharmacological (pharmacodynamics and pharmacokinetic) properties, including bioavailability, distribution, metabolism and toxicity. Hence, by combining the data obtained from all these techniques, we were able to obtain a comprehensive understanding of the mechanism of action of the compounds against NUPR1. This bottom-up approach enabled us to design drugs targeting IDPs, such as NUPR1, effectively.

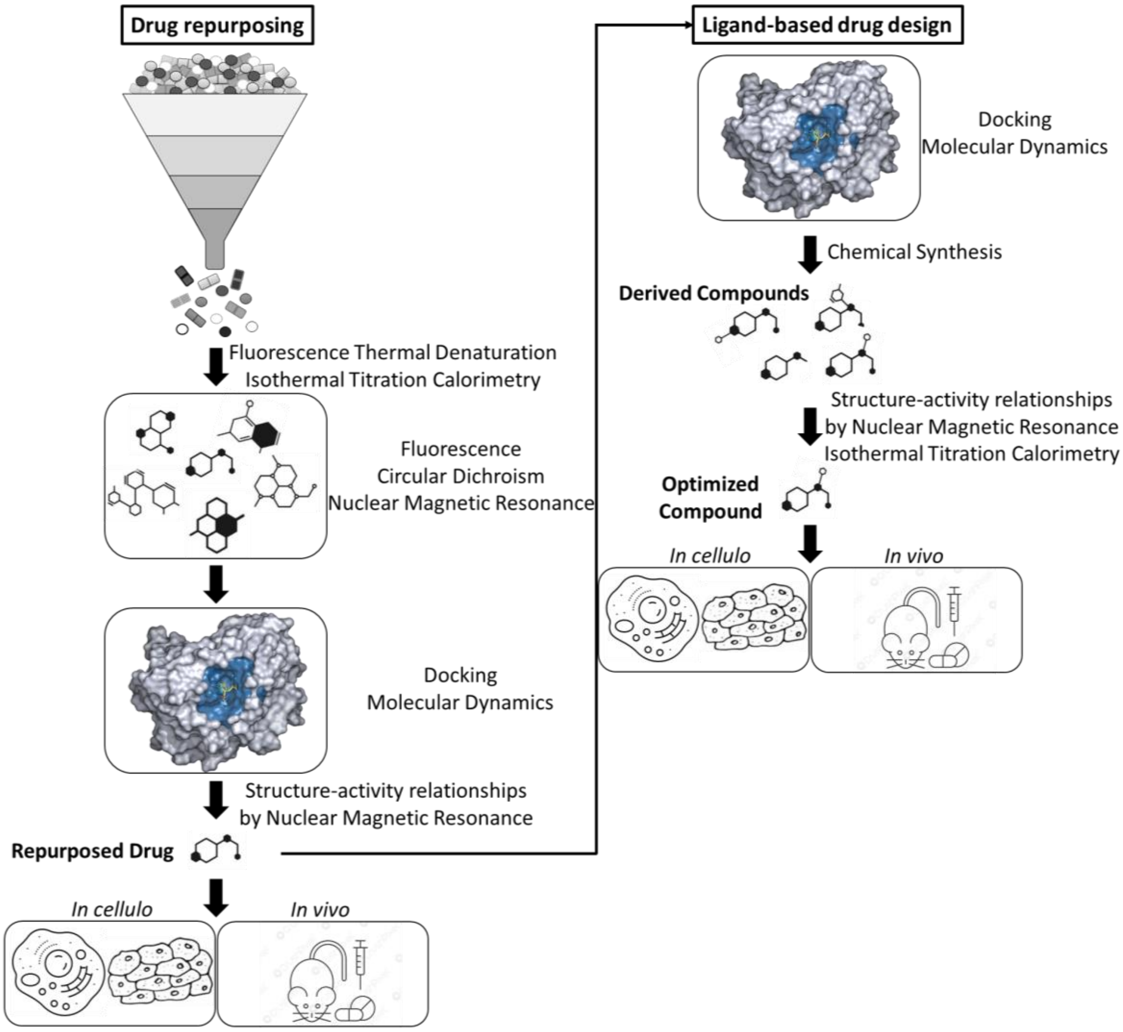
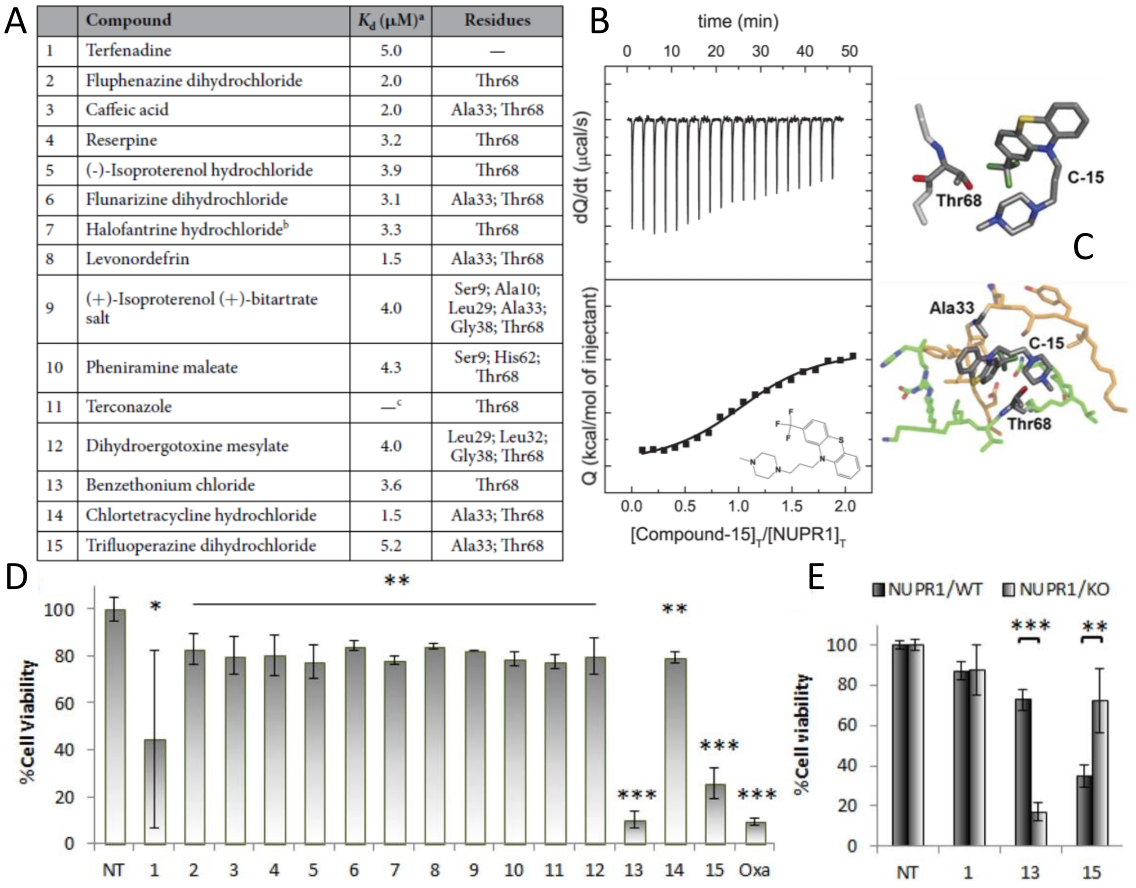

Drug repurposing
Fluorescence Thermal Denaturation
Isothermal Titration Calorimetry
Fluorescence
Circular Dichroism
Nuclear Magnetic Resonance
Docking
Molecular Dynamics
Structure-activity relationships
by Nuclear Magnetic Resonance
Repurposed Drug
In cellulo
In vivo
Ligand-based drug design
Docking
Molecular Dynamics
Chemical Synthesis
Derived Compounds
Structure-activity relationships
by Nuclear Magnetic Resonance
Isothermal Titration Calorimetry
Optimized
Compound
In cellulo
In vivo

Figure 67. Multi-disciplinary approach to drug discovery.

In the initial stage of our research, we focused on characterizing the interactions between NUPR1 and a selection of potential ligands. To accomplish this selection, we performed a screening of compounds using a chemical library of 1,120 FDA-approved drugs, known as

Prestwick Chemical Library. The screening process involved fluorescence thermal denaturation, wherein we identified compounds that induced the most significant shifts or alterations in the protein thermal denaturation traces. This initial screening provided valuable insights into potential ligands for NUPR1. Subsequently, we carried out a comprehensive four-step strategy, combining experimental and computational methods. First, we employed ITC to determine the thermodynamic parameters associated with the binding of compounds, selected after the screening, to NUPR1. This step allowed us to quantify the binding affinity of each compound. In parallel, we performed various spectroscopic techniques such as fluorescence, CD and NMR to obtain spectroscopic knowledge. Using this information, we performed molecular dynamics (MD) simulations to generate an ensemble of NUPR1 conformations, capturing its dynamic nature. Next, we employed this ensemble of conformations to perform docking simulations with the selected compounds, thereby predicting their binding modes and interactions with NUPR1. To further validate our findings, we conducted structure-activity relationship (SAR) assays using NMR techniques. Specifically, we examined the complexes formed between NUPR1 and compounds, focusing on detecting residues within NUPR1 that were affected by the presence of the selected compounds. This analysis allowed us to establish the SAR by identifying the specific residues influenced by the binding process. Remarkably, our blind strategy, which incorporated SAR-NMR and MD simulations, consistently yielded comparable results. The residues affected by binding in NUPR1 closely corresponded to those identified through our computational simulations. Therefore, this convergence served as a strong validation of our approach. Notably, the dissociation constants for the compounds determined by ITC were within the micromolar range, analogous to those observed for natural binding partners of NUPR1 (294). This further confirmed the relevance and applicability of our findings in the context of NUPR1 inhibition. Overall, our multi-disciplinary approach provided valuable insights into the interaction between NUPR1 and potential ligands, advancing our understanding towards the development of inhibitors targeting this protein.

At the end of this first study, 15 compounds were identified as the most promising candidates, and among them, Trifluoperazine (TFP), which was the Compound-15 in the study, was identified as the compound exhibiting high affinity for NUPR1 (K_d = 5.2 µM) and the ability to reduce cell viability in a NUPR1-dependent manner. In addition, TFP displayed selective binding to the predicted protein residues located in the identified “hot spots” (**Figure 68**) (180).



	Compound	K_d (μM)[a]	Residues
1	Terfenadine	5.0	—
2	Fluphenazine dihydrochloride	2.0	Thr68
3	Caffeic acid	2.0	Ala33; Thr68
4	Reserpine	3.2	Thr68
5	(-)-Isoproterenol hydrochloride	3.9	Thr68
6	Flunarizine dihydrochloride	3.1	Ala33; Thr68
7	Halofantrine hydrochloride[b]	3.3	Thr68
8	Levonordefrin	1.5	Ala33; Thr68
9	(+)-Isoproterenol (+)-bitartrate salt	4.0	Ser9; Ala10; Leu29; Ala33; Gly38; Thr68
10	Pheniramine maleate	4.3	Ser9; His62; Thr68
11	Terconazole	—[c]	Thr68
12	Dihydroergotoxine mesylate	4.0	Leu29; Leu32; Gly38; Thr68
13	Benzethonium chloride	3.6	Thr68
14	Chlortetracycline hydrochloride	1.5	Ala33; Thr68
15	Trifluoperazine dihydrochloride	5.2	Ala33; Thr68

Figure 68. (A) Table containing the affinity interaction measurements of the 15 NUPR1-Compound complexes (K_d) obtained by ITC at 25 °C, and affected residues of the protein after the binding of compounds with NMR-cross-peak broadening ([a]Relative error is 20%. [b]The compound precipitated at the NMR concentrations employed. [c]Not determined at the conditions tested). (B) Calorimetric titration for TFP (Compound-15) in the interaction with NUPR1 obtained by ITC at 25 °C. Thermogram (upper panel) and binding isotherm (lower panel) are shown. Non-linear fit according to a model considering a single ligand binding site (continuous line) and the molecular structure of the TFP are shown. (C) Binding mode of TFP (Compound-15) with the side chain of Thr68-a portion of the protein main-chain (light gray) is also shown. A transient protein pocket obtained by MD after combining binding modes of TFP (Compound-15) with NUPR1 main chain portions including residues 27-40 (orange) and 62-72 (green). (D) MiaPaCa-2 and (E) primary murine cell lines genetically modified (NUPR1 KO clones: no expression of *nupr1* gene) comparing with NUPR1 WT clones, treated with 10 μM of compounds for 6 days. Error bars are SD from 3 independent measurements (*p*-value: * ≤ 0.1; ** ≤ 0.05; *** ≤ 0.0001). Figure adapted from (180).

Regarding the activity of TFP *in vivo,* employing human PDAC cells-derived xenografts implanted into immunodeficient mice, a near-complete cessation of tumor growth in the treated animals with 10 mg/kg TFP was observed (**Figure 69**). More specifically, after 2 weeks of treatment, tumor volumes showed an exponential growth in the control group (1530 ± 184 mm^3), while the tumor volume increased only 50% in mice treated with 5 mg/kg TFP (767 ± 196 mm^3). Notably, in the case of mice treated with 10 mg/kg TFP, the tumor growth was rapidly, and almost completely, halted (558 ± 152 mm^3), even after the 4 weeks of the treatment (180).

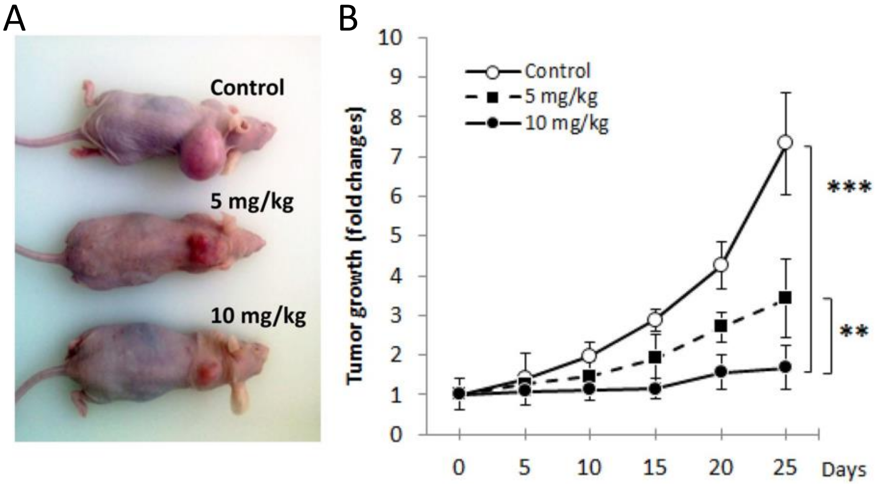
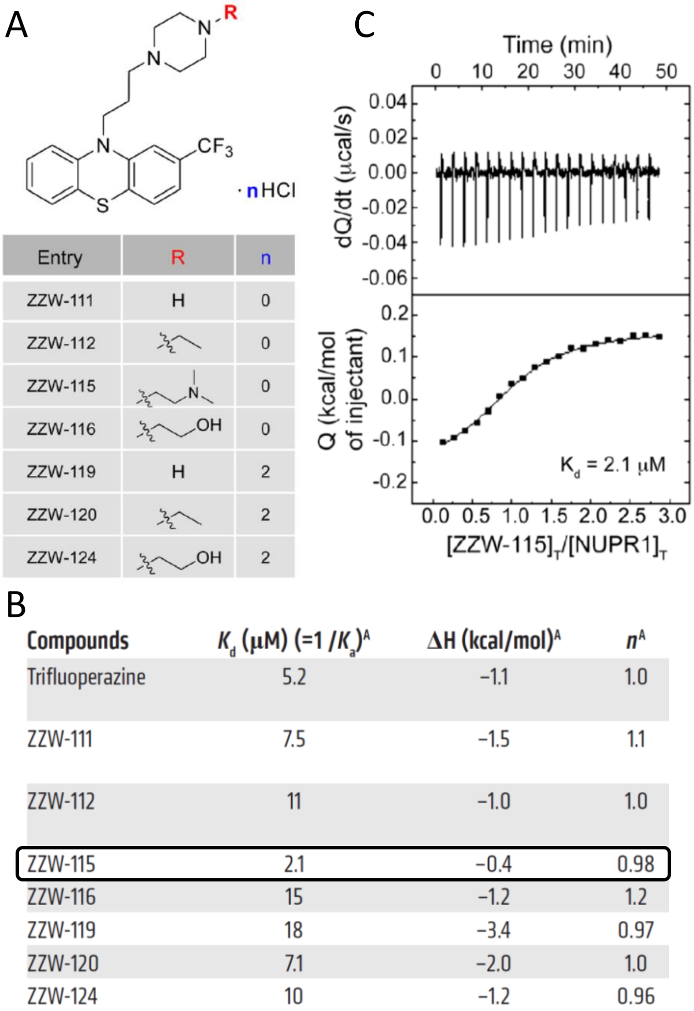


Figure 69. In human pancreatic tumors xenografts, TFP was able to stop the tumor growth. (A) Representative images of tumors after 4 weeks of daily intraperitoneally (i.p.) treatment with 0, 5 or 10 mg/kg TFP. The tumor size of the control group showed an exponential growth, whereas the tumors of mice treated with TFP remained constant. (B) The change in tumor growth was evaluated for each group of mice with 0, 5 or 10 mg/kg TFP during the 4 weeks of daily treatment. Analysis of variance for repeated measurements revealed a statistically significant treatment effect (*p*-value: ** ≤ 0.05; *** ≤ 0.01) when compared to the control group. Figure from (180).

This achievement signified a noteworthy case in drug repurposing and served as a pioneering attempt for an IDP (180). It is worth noting that these efforts in identifying and developing drugs against NUPR1 represent one of the first successful cases in the search for drugs against IDPs, a challenging endeavor that has been considered a technological landmark according by World Economic Forum's Forecast of 2019's Top Emerging Technologies (329).

Regrettably, the required TFP doses also induced undesirable neurological side-effects, such as pronounced lethargy and a hunched posture in the mice due to the fact that TFP is an antipsychotic agent primarily used to treat schizophrenia (330). Therefore, our efforts focused on enhancing the anticancer efficacy of TFP while minimizing its neurological side effects. To achieve this, we employed an *in silico* ligand-based design approach, combining MD simulations and docking techniques. This computational analysis guided the organic synthesis of novel TFP-derived compounds (ZZW compounds). The newly synthesized compounds were subjected to comprehensive biophysical analyses, including ITC and SAR-NMR, ultimately leading to the identification of a compound (ZZW-115), which was the one with the lowest K_d (**Figure 70**) (331).



Entry	R	n
ZZW-111	H	0
ZZW-112		0
ZZW-115	N	0
ZZW-116	OH	0
ZZW-119	H	2
ZZW-120		2
ZZW-124	OH	2

Compounds	K_d (µM) (=1 /K_a)^A	ΔH (kcal/mol)^A	n^A
Trifluoperazine	5.2	−1.1	1.0
ZZW-111	7.5	−1.5	1.1
ZZW-112	11	−1.0	1.0
ZZW-115	2.1	−0.4	0.98
ZZW-116	15	−1.2	1.2
ZZW-119	18	−3.4	0.97
ZZW-120	7.1	−2.0	1.0
ZZW-124	10	−1.2	0.96

Figure 70. Synthesis of compounds and characterization of the binding of NUPR1 to the compounds. (A) Structure of the synthesized TFP-derived compounds, called ZZW-compounds. (B) Thermodynamic parameters of the binding reactions of NUPR1 with the compounds. ZZW-115 was the TFP-derived compound with the highest affinity for NUPR1 (K_d = 2.1 µM). (C) Calorimetric titration of the NUPR1-ZZW-115 complex obtained by ITC. On the top: thermogram (thermal power as a function of time); and on the bottom: binding isotherm (ligand-normalized heat effect as a function of molar ratio). Figure adapted from (331).

Furthermore, cell-based experiments demonstrated a significant increase in the anticancer activity of ZZW-115 on PDAC-derived cells. Importantly, it was confirmed that this compound displayed its anticancer activity via specifically targeting NUPR1, since that *nupr1* KO clones were significantly more resistant to ZZW-115-treatment than *nupr1* WT clones (these clones had *nupr1* gene inactivated by a CRISPR/Cas9 approach). Notably, this TFP-derived compound also exhibited efficacy against other cancer cell lines (331). At the cellular level, lactate dehydrogenase (LDH) release, caspase-3/7 activity and ATP content were measured, observing that ZZW-115 was able to increase LDH release and caspase-3/7 activity, and to reduce ATP content, in a dose-dependent manner. Therefore, these results revealed that ZZW-115 compound simultaneously induced cell death through necroptotic (LDH assay) and apoptotic (caspase-3/7 assay) mechanisms by increasing LDH release and caspase-3/7 activity in a dose-dependent manner, as well as producing the disruption of the mitochondrial function by reducing ATP content in cells (**Figure 71**) (331).

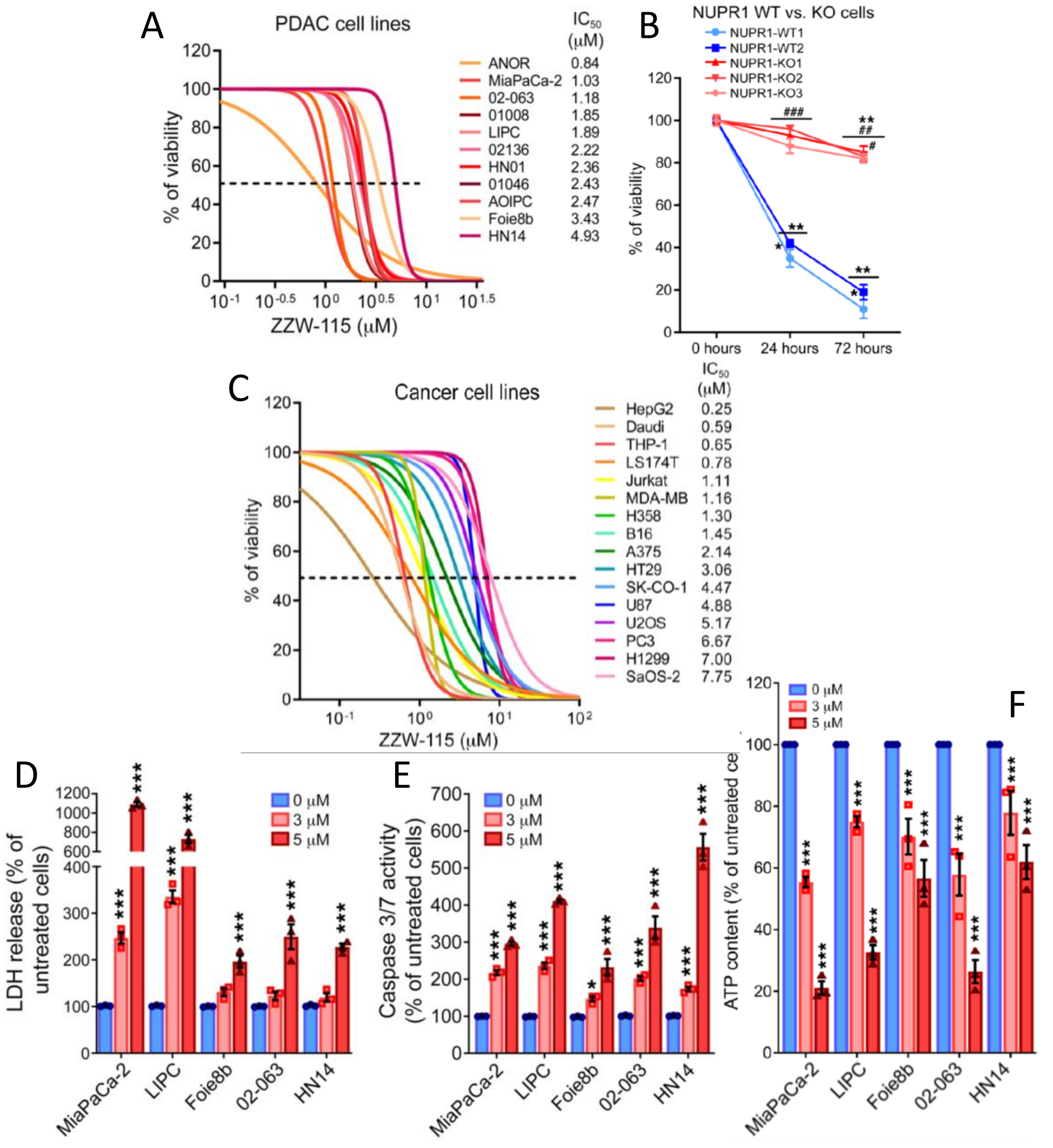


Figure 71. ZZW-115 compound showed antitumoral effects and induced cell death by necrosis and apoptosis *in vitro* by inhibiting NUPR1. (A) Viability of PDAC cancer cell lines, (B) NUPR1 WT or KO clones, and (C) different cancer cell lines after the treatment with increasing concentrations of ZZW-115 for 72 h. (D) LDH release, (E) caspase-3/7 activity and (F) ATP content measurement on PDAC cell lines incubated with 0, 3 or 5 µM of ZZW-115 for 24 h. Statistical significance was $*p < 0.05$ and $***p < 0.001$ compared with untreated group (two-way analysis of variance (ANOVA), Bonferroni's post hoc test). Data represent mean ± SEM, n=3. Figure adapted from (331).

Excitingly, according to *in vivo* assays involving xenograft pancreatic tumors in mice, ZZW-115 administration resulted in complete dose-dependent regression of the tumors with no observed side effects (**Figure 72**) (331). Importantly, ZZW-115 not only showed tumor regression, but also tumor elimination in the 5 mg/kg-dose group after 30 days of treatment. These results were more encouraging considering that they did not show neurological side effects, such as those showed by the TFP.

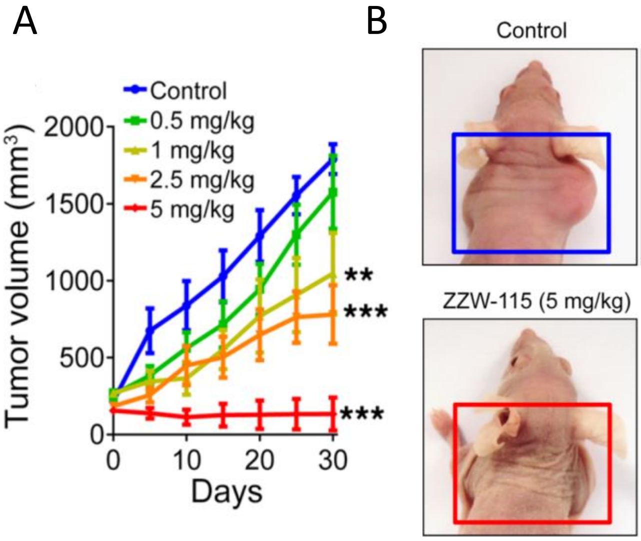


Figure 72. ZZW-115 had a strong antitumoral effect *in vivo*. (A) NMRI-Foxn1nu/Foxn1nu mice xenografted with MiaPaCa-2 cells were separated into 5 groups of 6 mice, which were daily i.p. treated (control group: 0.5% DMSO dissolved in 1X PBS; the other 4 groups: 5, 2.5, 1.0 or 0.5 mg/kg ZZW-115) for 30 days. Tumor volumes were measured every 5 days, and then the mean of the volume of each treatment was calculated. Statistical analysis was calculated using one-way ANOVA and Tukey's post hoc test (**$p < 0.01$ and ***$p < 0.001$). (B) Representative images of mice from both the control group and the 5 mg/kg ZZW-115 group. Figure adapted from (331).

From the previously mentioned results, ZZW-115 constituted a promising drug candidate for PDAC due to its unique molecular mechanism at that moment of the research, since it produced cell death by combining induction of necroptosis and apoptosis with mitochondrial metabolism failure. However, it was observed that ZZW-115 could bind to the potassium ion channel encoded by the human Ether-à-go-go-Related Gene (hERG), showing an inhibitory effect on this channel, thus increasing risk of cardiotoxicity (332,333). Thus, after the binding of ZZW-115 to hERG channel, it blocked the fast delayed rectifier current (I_{Kr}) of the heart, which induced a prolongation of the QT interval in the cardiac action potential. As a result, this QT interval prolongation increased the risk of developing a specific type of ventricular arrhythmia, called *torsades de pointes*, which could result in cardiac death (334,335). Therefore, identifying and developing novel NUPR1 inhibitors without compromising cardiovascular safety, i.e., safer and more selective for NUPR1, is absolutely needed to improve the treatment of PDAC.

2. OBJECTIVES

To develop new effective therapeutic agents against NUPR1 protein, associated to different pathologies such as PC and a huge amount of different kind of human cancers, and other pathologies including pancreatitis, inflammatory and neurological conditions, and diabetes.

To achieve this purpose, firstly it was required: 1/ To perform a functional characterization of the protein and study various properties of NUPR1, including its role in binding to potential biological partners. 2/ To identify compounds that interact with and inhibit NUPR1 through experimental molecular screening *in vitro*, followed by validation through *in vitro* techniques, cell cultures, and animal model assays.

The following general objectives were stablished in Chapter II:

1. **To conduct a comprehensive protein characterization** to examine various aspects of NUPR1 and its interactions with potential biological partners.
2. **To identify novel antineoplastic compounds** targeting the NUPR1 protein, which is relevant in the context of PC.
3. **To investigate the effects of the selected compounds** on the NUPR1 protein using *in vitro* and *in vivo* systems.

Regarding the first objective, "*To conduct a comprehensive protein characterization to examine various aspects of NUPR1 and its interactions with potential biological partners*", the following specific goals were proposed:

- To overexpress the NUPR1 protein in *E. coli* culture and purify the recombinant protein by affinity chromatography.
- To analyze the interaction between NUPR1 and Impα as a potential biological partner.
- To assess the role of the NLS sequence of NUPR1 for its nuclear translocation mediated by Impα.
- To investigate the autoinhibitory role of the IBB domain of Impα in NUPR1 recognition.
- To examine the three previous objectives in NUPR1L, the NUPR1 paralogue.
- To analyze the inhibitory activity of ZZW-115 for the interaction between NUPR1 and Impα.

Regarding the second objective, "*To identify novel antineoplastic compounds targeting the NUPR1 protein, which is relevant in the context of PC*", the following specific aims were proposed:

- To design and conduct a high-throughput screening (HTS) of compounds *in vitro* from a chemical library against NUPR1 as the target molecule. This aims to identify candidate compounds based on ligand-induced protein stabilization against thermal denaturation.
- To verify the interactions and determine the binding affinity between NUPR1 and the selected compounds in the screening.

Regarding the third objective, "*To investigate the effects of the selected compounds on the NUPR1 protein using in vitro and in vivo systems*", the following specific aims were proposed:

- To evaluate the antitumoral effects of the selected compounds *in vitro* using various cancer cell lines.
- To conduct *in vitro* biochemical assays to elucidate the mechanisms of action of the most promising compounds in inhibiting NUPR1.
- To analyze the antitumoral effect of the most promising compound *in vivo* in a mice animal model.

3. MATERIALS AND METHODS

3.1. Plasmid constructions

The expression vector pQE-30 (Qiagen) containing the human NUPR1 sequence and cloned in the *Bam*HI-*Hind*III site of the vector was used to generate a fusion protein with an N-terminal His-tag (MRGSHHHHHHGS) to the human NUPR1 sequence (**Figure 73**).

Positions of elements in bases	**pQE-30**
Vector size (bp)	3461
Start of numbering at *Xho*I (CTCGAG)	1–6
T5 promoter/lac operator element	7–87
T5 transcription start	61
6xHis-tag coding sequence	127–144
Multiple cloning site	145–192
Lambda t_0 transcriptional termination region	208–302
rrnB T1 transcriptional termination region	1064–1162
ColE1 origin of replication	1638
β-lactamase coding sequence	3256–2396

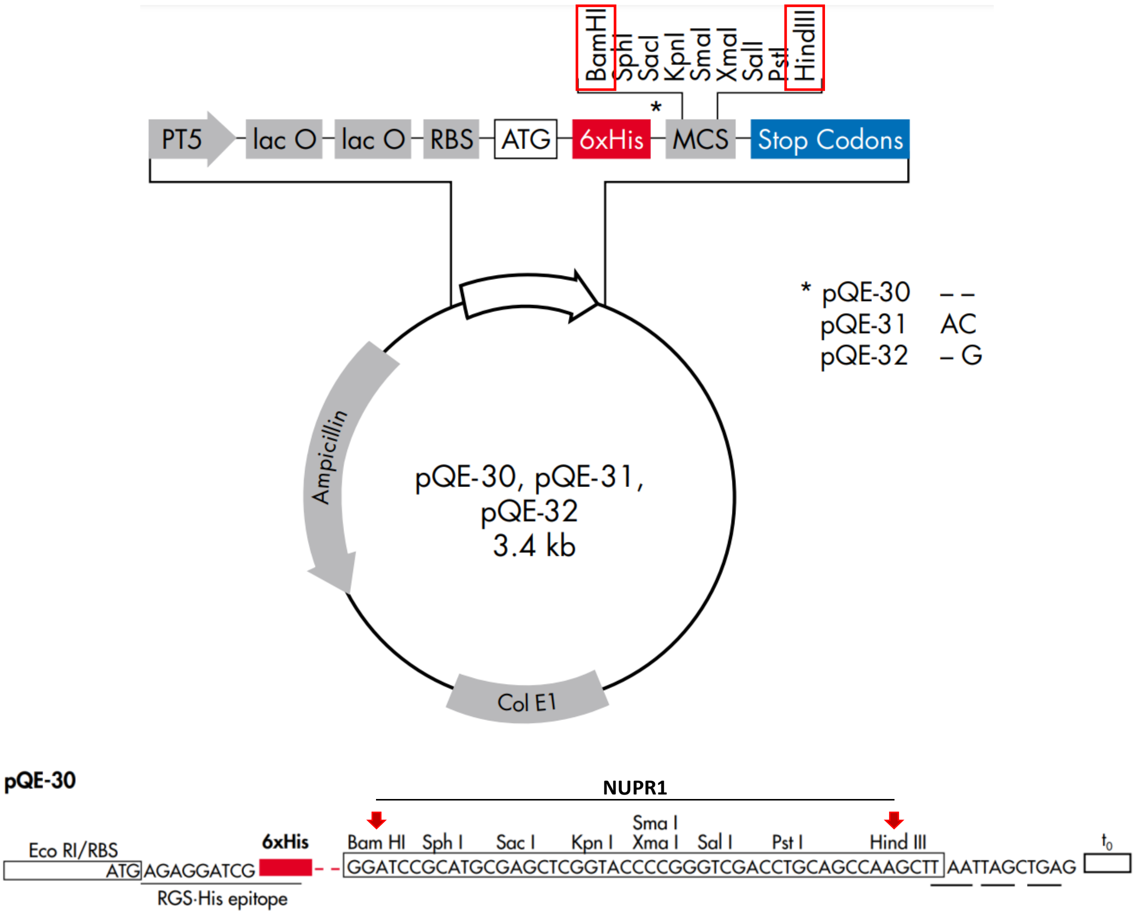

BamHI
SphI
SacI
KpnI
SmaI
XmaI
SalI
PstI
HindIII
*
PT5
lac O
lac O
RBS
ATG
6xHis
MCS
Stop Codons
* pQE-30 – –
pQE-31 AC
pQE-32 – G
Ampicillin
pQE-30, pQE-31, pQE-32
3.4 kb
Col E1
pQE-30
NUPR1
Eco RI/RBS
ATG AGAGGATCG
6xHis
Bam HI
Sph I
Sac I
Kpn I
Sma I
Xma I
Sal I
Pst I
Hind III
GGATCCGCATGCGAGCTCGGTACCCCGGGTCGACCTGCAGCCAAGCTT AATTAGCTGAG
t_0
RGS·His epitope

Figure 73. NUPR1 insert and detail of the expression vector pQE-30.

A codon-optimized vector comprising residues 1-520 of Impα3 with a His-tag at the N terminus was synthesized by NZYTech and cloned into the pHTP1 vector with Kanamycin (Kan) resistance.

ΔImpα3, i.e., Impα3 lacking the autoinhibitory IBB domain (residues 64-521), was His-tagged and codon optimized for BL21 (DE3) *E. coli* expression, synthesized and cloned into a pET15b vector at *Nde*I and *Eco*RI sites.

NUPR1L was synthetized in a GenScript *E. coli* expression vector, His-tagged and codon-optimized for its used in BL21 (DE3) *E. coli* strains, and subcloned in a pET-30 vector with Kan resistance.

3.2. Protein expressions and purifications

The expression and purification of the NUPR1 protein were carried out in a very similar way to that performed with the proBFT-3 protein (*see Chapter I*). For this reason, these processes are described in less detail in this section, emphasizing only the differences in methods and materials.

NUPR1 was also overexpressed by transforming Star BL21 *E. coli* strains which were grown in LB medium at 37 °C overnight, but in this case the antibiotic used was Ampicillin (Amp) (Sigma-Aldrich) at 100 µg/mL. Positive colonies were those selected due to the Amp resistance acquired by plasmid incorporation. Negative control consisted of non-transformed cells spread in a LB/Amp plate. Similarly to proBFT-3, 250 mL small-scale BL21 *E. coli* cell cultures were grown in LB/Amp at 37 °C overnight and they were used as pre-cultures to inoculate large-scale 6 L cultures of LB/Amp (100 µg/mL), which were incubated under the same conditions until reaching an $OD_{600} \approx 0.7$-0.8. Protein expression was induced with 1 mM IPTG at 18 °C overnight. Then, cells were harvested in the same way than in Chapter I. NUPR1 transformed cells were resuspended in 300 mM NaCl, 50 mM sodium phosphate, pH 7.4. The cell lysate was also obtained by sonication and cell debris was removed by serial centrifugations. Supernatants were clarified by 0.45 µm-pore filtration and then subjected to IMAC purification in an ÄKTA FPLC system. A cobalt HiTrap TALON column was used with an imidazole elution gradient in 150 mM NaCl, 50 mM sodium phosphate, pH 7.4. Purification procedure monitoring and data storage was developed by the UNICORN control software. Purity at each step was evaluated by SDS-PAGE and pure protein fractions were pooled and dialyzed to remove imidazole in 50 mM Tris-HCl pH 8. Final protein concentration was quantified by UV absorbance by using the theoretical extinction coefficient ε at 280 nm (NUPR1: 2,980 M^{-1} cm^{-1}) and potential DNA contamination in protein samples was always evaluated.

Impα3, ΔImpα3 and NUPR1L proteins were obtained directly in pure form after their expression and purification by collaborators of this project. The following are brief details of the steps required for these purifications.

Impα3 and ΔImpα3 expressions were performed in similar ways in BL21 *E. coli* strains in LB medium at 37 °C overnight, after induction using 1 mM IPTG when OD_{600} ≈ 0.6-1.0. The purification process was similar to that detailed for NUPR1, with the exception of the final polish purification step, which was performed using a Superdex G200 16/60 column in 200 mM NaCl, 50 mM Tris-HCl, pH 8.0, in an ÄKTA FPLC system.

NUPR1L expression was performed in BL21 *E. coli* strains in LB medium at 37 °C overnight after induction using 0.8 mM IPTG when OD_{600} ≈ 0.4-0.9. After the induction, temperature was decreased to 15 °C for 15-16 h, followed by several sonication steps in which NUPR1L was not in the clarified lysate, but in the precipitate. This precipitate was diluted in 8 M urea, 500 mM NaCl, 0.1% Triton X-100, 1 mM β-mercaptoethanol, 5 mM imidazole, 20 mM Tris-HCl, pH 8.0, supplemented with a protease inhibitor cocktail, to perform another cycle of sonication before applying an IMAC purification with a Ni-resin column. On-column refolding was performed in 500 mM NaCl, 1 mM β-mercaptoethanol, 20 mM imidazole, 20 mM Tris-HCl pH 8.0, and the eluted NUPR1L was finally obtained in 500 mM NaCl, 1 mM β-mercaptoethanol, 100 mM sodium phosphate, pH 7.0.

3.3. Design, prediction and synthesis of NLS-NUPR1 and NLS-NUPR1L peptides

The peptides corresponded to the NLS region of NUPR1 were synthesized by NZYTech with a purity of 95%, and each sequence was named with a specific name indicated within parentheses (**Table 13**). To prevent fraying effects, all peptides were acetylated at the N-terminus and amidated at the C-terminus. Of note, since the WT NLS has no tyrosines, a tyrosine residue at the N-terminus for absorbance measurements was introduced. In total, eight peptides with different mutations were synthesized for different purposes:

- Mutations at positions Lys65 and Lys69 to Ala were used to study their importance in the binding to both importins.
- Thr68 was mutated to the glutamic T68E peptide to have a phosphomimetic at this position.
- Double mutants were obtained by combining the Thr68 mutation with either of the other two mutations, and a triple mutant was designed with all three mutations.

- A phosphorylated peptide at position Thr68 (pT68 peptide) was synthesized to study the effects of this specific PTM.

Peptide
YT^{54}NRPSPGGHERKLVTKLQNSE (WT)
YTNRPSPGGHER**A**LVTKLQNSE (K65A)
YTNRPSPGGHERKLVT**A**LQNSE (K69A)
YTNRPSPGGHERKLV**E**KLQNSE (T68E)
YTNRPSPGGHER**A**LV**E**KLQNSE (K65AT68E)
YTNRPSPGGHERKLV**EA**LQNSE (T68EK69A)
YTNRPSPGGHER**A**LV**EA**LQNSE (K65AT68EK69A)
YTNRPSPGGHERKLV**pT**KLQNSE (pT68)

Table 13. Detail of the sequences of the NLS-NUPR1 peptides. Mutations with respect to the WT sequence are showed in bold, and the phospho-threonine at position 68 of the last peptide is indicated with a “pT”.

The peptides corresponding to the NLS region of NUPR1L were also synthesized by NZYTech with a purity of 95%. To identify the NLS region of NUPR1L, the entire sequence of NUPR1L was subjected to NLS prediction using the following web server: http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi. The predicted region with the highest score was found to comprise residues Gly46 to Gln74. Subsequently, the peptide was strategically designed for enhanced solubility, comprising residues Arg51 to Gln74 (**Table 14**). To prevent fraying effects, the N-terminus was acetylated, and the C-terminus was amidated.

Peptide
RTRREQALRTNWPAPGGHERKVAQ

Table 14. Detail of the sequence of the NLS-NUPR1L peptide.

3.4. ITC assays

A high-sensitivity isothermal titration calorimeter Auto-iTC200 (MicroCal, Malvern-Panalytical) was used to study the different interactions studied in Chapter II in the same way that was explained in the Chapter I. For this reason, this technique is described in less detail in this section, emphasizing only the differences in protein concentrations, and buffer and temperature conditions.

All the assays related to the first aim of Chapter II, i.e., "*To conduct a comprehensive protein characterization to examine various aspects of NUPR1 and its interactions with potential biological partners*", were performed in 50 mM Tris-HCl, pH 8.0, at a constant temperature of 25 °C, and in the following specific conditions:

- For ITC assays to evaluate the interaction between NUPR1 and Impα3 or ΔImpα3, 5-10 µM NUPR1 were loaded into the calorimetric cell and 100-110 µM Impα3 or ΔImpα3 into the syringe.
- For ITC assays to evaluate the interaction between NLS-NUPR1 peptides and Impα3 or ΔImpα3, 10-20 µM Impα3 or ΔImpα3 were loaded into the calorimetric cell and 150-300 µM NLS-NUPR1 peptides into the syringe.
- For ITC assays to evaluate the interaction between NLS-NUPR1L peptides and Impα3 or ΔImpα3, 10-20 µM Impα3 or ΔImpα3 were loaded into the calorimetric cell and 150-300 µM NLS-NUPR1L peptides into the syringe.
- For ITC assays to evaluate the interaction between NUPR1/Impα3/ZZW-115 compound, 5-10 µM NUPR1 were loaded into the calorimetric cell and 100-110 µM Impα3 into the syringe. A concentration of 100 µM ZZW-115 compound was employed in the calorimetric cell.

All the assays related to the second aim of Chapter II, i.e., "*To identify novel antineoplastic compounds targeting the NUPR1 protein, which is relevant in the context of PC*", were performed in 50 mM sodium phosphate, pH 7.0, at a constant temperature of 15 °C, and in the following specific conditions:

- For ITC assays to evaluate the interaction between NUPR1 and the candidate compounds, selected in the experimental screening, 20 µM NUPR1 was titrated in the calorimetric cell with 200 µM compound solutions from the syringe.

3.5. Chemical libraries

One commercially available chemical library was employed for the experimental molecular screening: HitFinder Chemical Library (Maybridge Chemical) containing 10,000 compounds from disparate sources, providing a huge chemical and pharmacological diversity. Compounds were dissolved in 100% DMSO at a final concentration of 10 mM, and they were arranged in 96-well plates preserved at -20 °C while not in use.

3.6. Experimental ligand screening: TSA assays

Potential ligands for NUPR1 were identified from the Maybridge Library following a HTS procedure based on TSA by DSF. This procedure was similar to that employed previously in Chapter I, but it presented some differences due to the fact that NUPR1 is an IDP, in contrast to the well folded BFT-3 protein.

Thermal denaturations of NUPR1 in the presence of compounds were performed in a Stratagene Mx3005P real-time qPCR thermal cycler (Agilent) using the extrinsic fluorescent ANS (Sigma Aldrich) as protein unfolding reporter. 100 µL of protein solutions containing 4 µM NUPR1, 100 µM ANS and 250 µM compound (2.5% residual final concentration of DMSO) in 150 mM NaCl, 50 mM sodium phosphate, pH 7.4 were dispensed into 96-well microplates (96-Well PCR Plate, Non-skirted from 4titude) and incubated at room temperature for 10-15 min before loading onto the microplate reader. The unfolding curves were registered, as in Chapter I, from 25 to 95 °C, at a 1 °C/min scan rate, but in this case following the increase in ANS emission fluorescence intensity (λ_{exc} = 330 and λ_{em}= 492 nm), which greatly increases when this probe binds to hydrophobic regions in the protein exposed to the solvent upon thermal unfolding.

Control experiments of NUPR1 samples with DMSO were routinely performed in each microplate. Hits were identified as those compounds shifting the temperature for maximal slope towards higher temperatures or those altering significantly the denaturation profile, compared to the internal controls in each microplate, thus inducing a stabilizing effect on NUPR1 and potentially capable of inhibiting any subsequent protein-NUPR1 interactions. Although NUPR1 is an IDP devoid of stable well-folded structure, the binding of some compounds could promote local folding on NUPR1, and thermal destabilization and unfolding of that structure could be observed in the assays.

3.7. Nomenclature of selected compounds in the *in vitro* experimental ligand screening

The chemical compounds selected in the experimental screening by TSA for Chapter II were also internally coded for a quick and easy identification.

The selected compounds from the Maybridge Chemical Library were named AJOX, with "X" being the compound identifier number from 1 to 45, i.e., AJO1, AJO2, AJO3…

3.8. Cell lines and cell cultures

The following human cancer cell lines were cultured under specific conditions:

- MiaPaCa-2: human pancreatic cancer cells.
- HepG2: human hepatocellular carcinoma cells.
- MDA-MB-231: human breast adenocarcinoma cells.
- A375: human melanoma cells.
- HT-29: human colorectal cancer cells.
- U87: human primary glioblastoma cells.
- U2OS: human bone osteosarcoma epithelial cells.
- H358: human lung carcinoma cells.
- PC-3: human prostate cancer cells.
- Primary human PDAC cells.
- Panc-1 clones: wild-type *nupr1* clones (NUPR1-WT) and knockout *nupr1* clones (NUPR1-KO), previously developed using CRISPR-Cas9 technology in Panc-1 cells (human pancreatic cancer cells) (305).

The culture media used were as follows: MiaPaCa-2, HepG2, MDA-MB-231, A375, HT-29, U87, U2OS and Panc-1 clone cells were grown in DMEM (Gibco, Invitrogen), supplemented with 10% FBS (Biosera). H358 and PC-3 cells were cultured in RPMI 1640 medium (Gibco, Invitrogen), supplemented with 10% FBS. Primary human PDAC cells were grown in serum-free DMEM/F12 medium, which was supplemented with: 1.22 g/L nicotinamide, 5 g/L glucose, 5% Nu-Serum IV, 0.5% ITS + Premix Universal Culture Supplement (insulin, human transferrin, and selenous acid), 1 µM dexamethasone, 10 ng/L cholera toxin, 50 nM 3,3',5-triiodo-L-thyronine, 25.2 mg/L bovine pituitary extract and 20 µg/L epidermal growth factor. All cells were incubated in a 5% CO_2 humidified incubator at 37 °C.

3.9. Cell viability assays

5,000 cells/well were plated in 96-well plates and allowed to attach by incubating for 24 h. Subsequently, cells were exposed to AJO compounds at a concentration range of 0-100 µM for 72 h at 37 °C. After the incubation period, cell viability was assessed by adding 10 µL PrestoBlue^{TM} Cell Viability Reagent (Thermo Scientific) to each well containing 100 µL of blank, control or treated cells in culture, and incubated for another 3 h at 37 °C. When added to media, the PrestoBlue reagent is rapidly taken up by cells. Briefly, the reducing environment within viable cells converts the non-toxic resazurin present in the PrestoBlue reagent to an intensely red-fluorescent dye. The changes in cell viability were detected by measuring fluorescence (with λ_{exc} and λ_{em} set at 570 nm and 610 nm, respectively) using the Tristar LB941 plate reader. Cell viability was normalized with respect to untreated cell rates. All samples and each experiment were performed in triplicate.

3.10. hERG channel binding assay

As previously indicated in the Introduction of Chapter II, a very important aspect for a drug to reach clinical trials in humans nowadays is that it must not bind to the potassium hERG channel.

The measurement of hERG potassium channel blocking propensity was performed using the Predictor™ hERG Fluorescence Polarization Assay kit (Thermo Fisher Scientific). To perform the assay, membrane fractions containing hERG channel protein were employed. Several concentrations of AJO14 compound from 0.1-80 µM were incubated with the membrane fractions and the resulting membrane polarization (mP) values were recorded. As a positive control, E-4031, a selective hERG potassium channel blocker, was used. The compounds were initially prepared at a 100X concentration in DMSO and then diluted to 4X (4% DMSO) in the assay buffer. Each well of the 386-well plate contained 5 µL of the assay buffer, and 10 µL of the hERG membrane samples were used as the negative control. Subsequently, 5 µL of the compounds were added to each well, resulting in a final concentration range of 0.1-80 µM. Next, 10 µL of 2X Predictor hERG membranes were dispensed into each well, followed by 5 µL of 4X Predictor hERG Tracer Red. To account for potential interference from compounds, each component was tested both in the absence and presence of 30 µM E-4031. The plates were incubated for a minimum of 7 h before measuring the fluorescence polarization. A BMG LABTECH device (PHERAstar) was used to measure the fluorescence polarization, with λ_{exc} and λ_{em} set at 540 and 573 nm, respectively.

3.11. LDH assay, caspase-3/7 activity assay and ATP production assay

MiaPaCa-2 cells were seeded at a density of 10,000 cells/well in 96-well plates. Then, plates were incubated to allow the cells to adhere for 24 h. After this period, cells were treated with increasing concentrations of AJO14 until 80 µM for 24, 48 and 72 h. All treatments were performed in triplicate.

To assess cellular responses after AJO14 treatment, three different commercial kits were used:

- LDH release: the release of LDH was measured using the CytoTox-ONE^{TM} Homogeneous Membrane Integrity Assay (Promega). This assay was a rapid fluorometric method used to determine cytotoxicity by estimating the non-viable cells present in a sample. The amount of LDH released from cells with damaged membrane provided an indication of cell death (necrosis-associated processes) (**Figure 74**). The changes in LDH release were detected by measuring fluorescence (with λ_{exc} and λ_{em} set at 570 nm and 610 nm, respectively) using the Tristar LB941 plate reader.

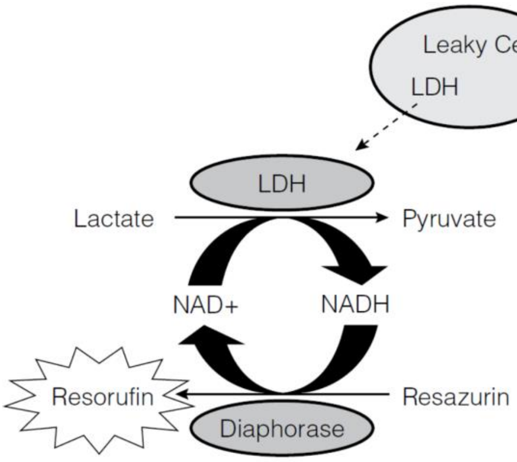


Figure 74. Principle of LDH release detection employed in the CytoTox-ONE^{TM} Homogeneous Membrane Integrity Assay. The release of LDH from damaged cells was measured by lactate, NAD$^+$, and resazurin as substrates in the presence of diaphorase. LDH catalyzed the conversion of lactate to pyruvate while reducing NAD$^+$ to NADH. Then, the NADH produced reduced the resazurin dye to its fluorescent product, called resorufin, which was proportional to the amount of LDH present in the sample.

- Caspase-3/7 activity: the activity of caspase-3/7, which are key enzymes involved in apoptosis (programmed cell death), was measured using the Caspase-Glo® 3/7 Assay (Promega), which was luminescent assay that detected the presence of active caspase-3 and caspase-7. The amount of caspase activities provided information on apoptotic cell death (**Figure 75**). The changes in the activity of caspase-3/7 were detected by measuring luminescence using the Tristar LB941 plate reader.

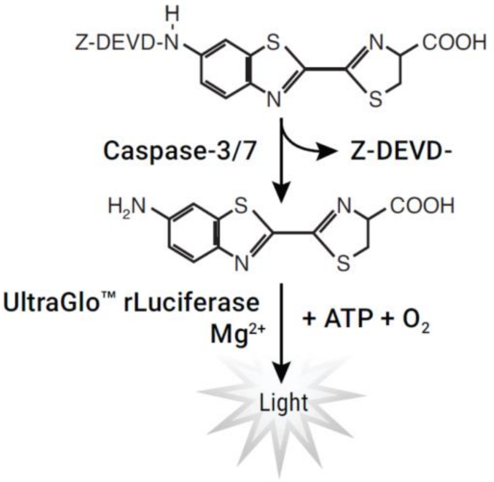

Z-DEVD-N
H
S
N
COOH
S
N
Caspase-3/7
Z-DEVD-
H2N
UltraGlo™ rLuciferase
Mg2+
+ ATP + O2
Light

Figure 75. Principle of caspase-3/7 activity detection employed in the Caspase-Glo® 3/7 Assay. After caspase cleavage of the proluciferin DEVD substrate, the substrate for luciferase (aminoluciferin) was released. In presence of luciferase and ATP, there was a generation of light, which was directly proportional to the amount of caspase activity in the sample.

- ATP production: the production of ATP, which is an indicator of cellular energy metabolism, was measured using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega), which was a luminescent assay that quantifies ATP levels in viable cells and served as an indicator of cell viability and metabolic activity (**Figure 76**). The changes in ATP production were detected by measuring luminescence using the Tristar LB941 plate reader.

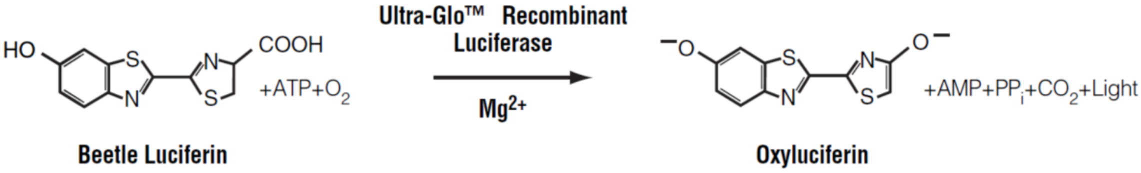

HO
COOH
+ATP+O2
Ultra-Glo™ Recombinant Luciferase
Mg2+
Beetle Luciferin
−O
O−
+AMP+PPi+CO2+Light
Oxyluciferin

Figure 76. Principle of luciferase reaction measured in the CellTiter-Glo® Luminescent Cell Viability Assay. Luciferase catalyzed the mono-oxygenation of luciferin in the presence of Mg^{2+}, ATP and O_2. Luciferase produced the incorporation of O_2 into luciferin, resulting in an excited intermediate, which released energy in the form of light. This reaction needed both Mg^{2+} and ATP as co-factors. The amount of ATP was directly proportional to the number of active viable cells present in a sample.

For LDH and caspase-3/7 activity assays, data obtained were normalized to cell number (see the crystal violet protocol explained below). In the case of ATP assay, cells were normalized to untreated cells.

3.12. Crystal violet

To normalize LDH and caspase-3/7 activity assays to the cell number, crystal violet assays were performed. Cells were plated in triplicate in 96-well plates and allowed to attach overnight. Subsequently, the cells were treated with increasing concentrations of AJO14 for 24, 48 and 72 h. After this incubation, culture medium was discarded, cells were washed with 1X PBS and then fixed using 1% glutaraldehyde. Then, cells were washed with 1X PBS and stained with 0.1% crystal violet in 70% methanol. The excess of crystal violet was removed, and cells were washed three times with 1X PBS. Finally, 1% SDS was added to the wells in order to solubilize the crystal violet and then the absorbance was measured at 590 nm using an Epoch™ Microplate Spectrophotometer.

3.13. Animals

Female Nude-Foxn1^{nu} mice at 4 weeks of age were obtained from Charles River Company and housed in the Experimental Animal House of the *Centre de Cancérologie de Marseille*, *pôle Luminy*. Mice were maintained in a specific pathogen-free environment and handled in accordance with laboratory animal care principles and ethical guidelines. All experimental procedures on animals were approved by the *Comité d'ethique de Marseille numéro 14* (C2EA-14). They were provided with an appropriate nutrition during the study. For the tumor inoculation, a total of 10^6 MiaPaCa-2 cells suspended in 50 µL of Matrigel (BD Pharmingen) were subcutaneously (s.c.) injected into 5-week-old nude mice. To initiate the treatment, once the tumor size reached 200 mm^3, mice were i.p. administered with daily doses of 0.5% DMSO in sunflower seed oil (vehicle) in the case of controls, and the additional groups with 5, 10, 20, or 50 mg/kg of AJO14. During this assay, mice were weighed as well as their tumor volumes were measured twice per week. Tumor volumes were estimated using the following formula: tumor volume = $(L \times W^2)/2$ (W represents the smaller dimension and L represents the larger dimension of the tumor). After 28 days of treatment, mice were euthanized.

3.14. Statistics

The statistical analyses were performed using different methods depending on the experimental design. The unpaired two-tailed Student t-test for comparisons between two groups, and either one-way ANOVA with Sidak correction or two-way ANOVA with Sidak correction for comparisons among multiple groups, were employed. The results were expressed as the mean ± SD of at least three independent experiments to ensure statistical robustness. A *p*-value of <0.05 was considered as statistically significant, indicating a meaningful difference between the compared groups or conditions.

4. RESULTS

4.1. Functional characterization of NUPR1 and its interaction with Impα3 as biological partner

4.1.1. NUPR1 was correctly purified

In the same manner as described in Chapter I, we conducted the necessary steps for the expression and purification of the NUPR1 protein to obtain it in a pure form, isolated from other proteins not of interest. Using SDS-PAGE gels, the purity of NUPR1 was analyzed after each purification. As depicted in **Figure 77**, at the end of all the purification steps a single band was consistently observed. Of note, the molecular weight (MW) of NUPR1 is 8.8 kDa, but sometimes in the case of very small proteins, the SDS-PAGE migration may not precisely correspond to their theoretical weight. To confirm the accurate identity of NUPR1, MALDI-TOF/TOF identification was performed. This analysis provided definitive confirmation that the protein purification was successfully executed.

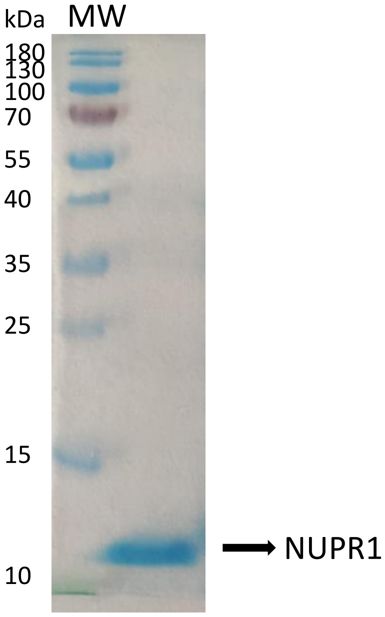


Figure 77. SDS-PAGE of pure NUPR1. MW: molecular weight protein ladder (PageRuler^{TM} Prestained Protein Ladder).

4.1.2. NUPR1 and Impα3 interacted *in vitro*

Throughout this Doctoral Thesis, as previously mentioned, our research group, in collaboration with our partners, identified 656 proteins capable of interacting with NUPR1. Among these proteins, over 30 were identified as components of the nuclear pore complex, including certain importins (292).

It is worth noting that although NUPR1 distributed throughout the cell, it contains a NLS enabling its translocation into the cell nucleus to perform its functions in eukaryotic cells. Despite being a small protein, its unfolded structure and considerable radius of gyration (282) suggest that it may rely on the importin system for proper nuclear translocation. Therefore, investigating the interaction between importins and NUPR1 is crucial for gaining insights into the mechanisms of NUPR1 nuclear translocation. Consequently, we undertook an *in vitro* evaluation of the interaction between NUPR1 and Impα3 using ITC (**Figure 78**). Of note, Impα3 was chosen for these studies due to its greater flexibility when compared with other importins, which enhances its ability to interact with cargo molecules, resulting in a wider range of conformations (327). Furthermore, Impα3 was readily expressed and purified. The results of the ITC assays revealed that NUPR1 displayed micromolar-range affinity for Impα3 (K_d = 1.4 µM) (**Table 15**), consistent with and similar to its binding affinity for other biomolecules (297,298).

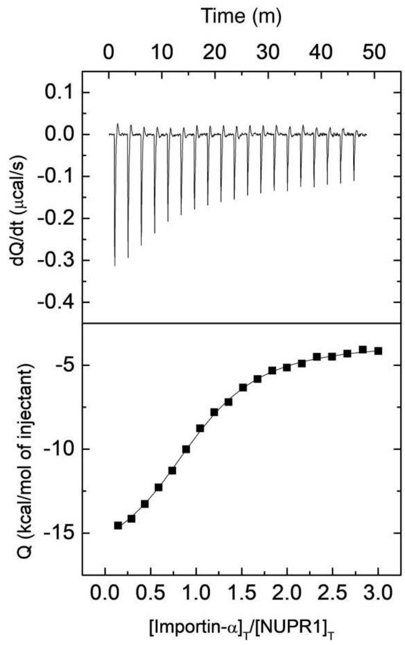


Figure 78. The interaction between NUPR1 and Impα3 was evaluated by ITC. Thermogram (upper panel; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherm (lower panel; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fit according to a model considering a single ligand binding site (continuous line) is shown.

	K_a (M^{-1})	**K_d (μM)**	**ΔH (kcal/mol)**	**ΔG (kcal/mol)**	**n**
Impα3/NUPR1	$6.9 \cdot 10^5$ [$6.5 \cdot 10^5$, $7.3 \cdot 10^5$]	1.4 [0.9, 1.9]	-13.7 [-14.2, -13.2]	-7.9 [-8.4, -7.4]	1.1 [0.9-1.3]

Table 15. Thermodynamic parameters for the Impα3/NUPR1 interaction determined by ITC at 25°C.

Based on these results, it was verified that NUPR1 was able to interact with Impα3 with a high affinity, so that the internalization of NUPR1 to the cell nucleus probably occurs by the mechanism of the classical nuclear protein import carried out by importins.

4.1.3. NUPR1 interacted also with ΔImpα3 *in vitro*

In the process of protein internalization into the nucleus by importins, the IBB domain of Impα3 is known to be important in modulating the assembly complex formation between importins and their cargos. Concretely, the IBB domain, which contains a large number of lysine residues, exerts an autoinhibitory effect, hampering or modulating the entrance of NLS peptides into the major NLS-binding region of Impα3 (328).

First, it was analyzed whether the absence of the IBB domain in the ΔImpα3 truncated protein affected the binding of NUPR1 by ITC (**Figure 79**).

The results provided a value for the K_d of 0.4 μM (**Table 16**), in comparison with the previously detailed K_d value of 1.4 μM for the intact Impα3. Therefore, the affinity of NUPR1 for Impα3 was higher in the absence of the IBB domain, i.e., NUPR1 had more affinity for ΔImpα3 than for Impα3, confirming that the IBB domain exerted an autoinhibitory effect, penalizing energetically the interaction with NUPR1.

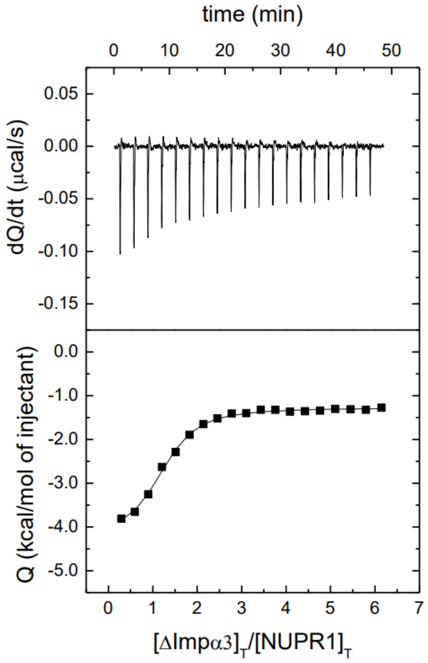


Figure 79. The interaction between NUPR1 and ΔImpα3 was evaluated by ITC. Thermogram (upper panel; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherm (lower panel; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fit according to a model considering a single ligand binding site (continuous line) is shown.

	K_a (M^{-1})	**K_d (μM)**	**ΔH (kcal/mol)**	**ΔG (kcal/mol)**	**n**
ΔImpα3/NUPR1	$2.5 \cdot 10^6$ [$2 \cdot 10^6$, $3 \cdot 10^6$]	0.4 [-0.1, 0.9]	-3.5 [-4.0, -3.0]	-8.7 [-9.2, -8.2]	0.9 [0.7-1.1]

Table 16. Thermodynamic parameters for the ΔImpα3/NUPR1 interaction determined by ITC at 25°C.

4.1.4. The NLS-NUPR1 peptides interacted with both Impα3 and ΔImpα3 *in vitro*

Once it was verified that NUPR1 interacted with Impα3 and ΔImpα3, and given the importance of both the NLS sequence in the binding between importins and cargos, and the fact that PTMs can induce IDPs to reach different structures at local levels thus modeling their biological functions, the following step was to analyze if the theoretically predicted NLS region of NUPR1 was capable of binding in isolation to Impα3 and ΔImpα3. In order to achieve this aim, we followed an approach based on studying the individual peptides from this region. In this approach, while the primary focus was to understand the interaction between the full-length

NUPR1 and several mutants with Impα3 and ΔImpα3, the use of NLS peptides to elucidate the specific binding mechanism with Impα3 was employed. The reason for adopting this method was due to the fact that, very often, mutations introduced at any place of the polypeptide length of NUPR1 resulted in a poor expression of the corresponding mutant, and in some cases, certain mutations led to no expression at all. Therefore, ITC experiments to measure the binding affinity between NLS-NUPR1 peptides and both Impα3 and ΔImpα3 were carried out (**Figure 80, and Table 17 and Table 18**).

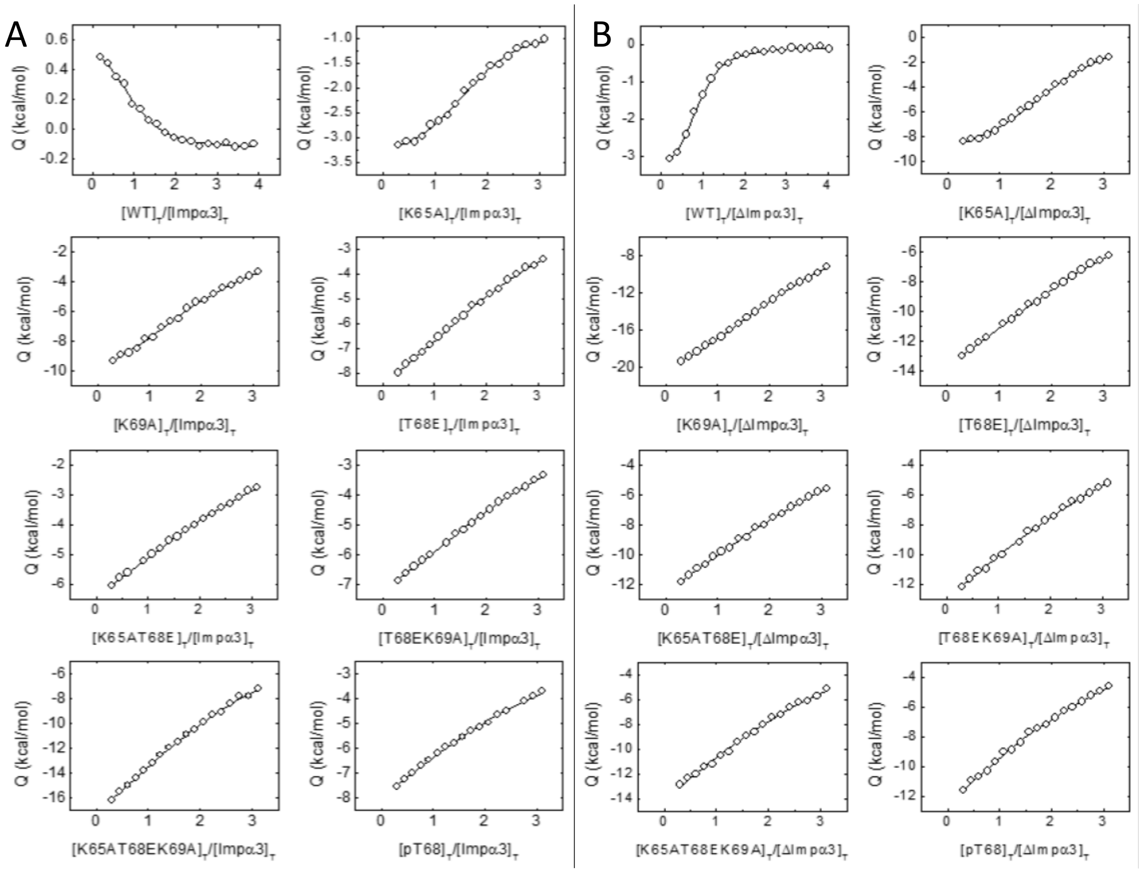


Figure 80. The interaction of the WT and mutant NLS-NUPR1 peptides with (A) Impα3 and (B) ΔImpα3 was measured by ITC. Thermograms (upper panels; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherms (lower panels; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fits according to a model considering a single ligand binding site (continuous line) are shown.

Peptide	Impα3				
	K_a (M^{-1})	K_d (μM)	ΔH (kcal/mol)	ΔG (kcal/mol)	n
WT	$5.9\cdot10^{5}$ $[4.2\cdot10^{5}, 7.3\cdot10^{5}]$	1.7 [1.3, 2.1]	0.8 [0.3, 1.3]	-7.9 [-8.4, -7.4]	0.9 [0.7, 1.1]
K65A	$2.6\cdot10^{5}$ $[1.9\cdot10^{5}, 3.4\cdot10^{5}]$	3.9 [3.5, 4.3]	-2.8 [-3.3, -2.3]	-7.4 [-7.9, -6.9]	1.4 [1.2, 1.6]
K69A	$9.1\cdot10^{4}$ $[7.5\cdot10^{4}, 10^{5}]$	11.0 [10.5, 11.5]	-10.8 [-11.3, -10.3]	-6.8 [-7.3, -6.3]	1.3 [1.1, 1.5]
T68E	$4.5\cdot10^{4}$ $[3.5\cdot10^{4}, 5.5\cdot10^{4}]$	22.0 [21.5, 22.5]	-11.1 [-11.6, -10.6]	-6.4 [-6.9, -5.9]	(1)
K65AT68E	$4.8\cdot10^{4}$ $[3.9\cdot10^{4}, 5.7\cdot10^{4}]$	21.0 [20.5, 21.5]	-7.8 [-8.3, -7.3]	-6.4 [-6.9, -5.9]	(1)
T68EK69A	$5.9\cdot10^{4}$ $[5\cdot10^{4}, 6.8\cdot10^{4}]$	17.0 [16.5, 17.5]	-7.5 [-8.0, -7.0]	-6.5 [-7.0, -6.0]	(1)
K65AT68EK69A	$3.7\cdot10^{4}$ $[3\cdot10^{4}, 4.4\cdot10^{4}]$	27.0 [26.5, 27.5]	-16.3 [-16.8, -15.8]	-6.2 [-6.7, -5.7]	(1)
pT68	$3.7\cdot10^{4}$ $[3\cdot10^{4}, 4.4\cdot10^{4}]$	27.0 [26.5, 27.5]	-14.8 [-15.3, -14.3]	-6.2 [-6.7, -5.7]	(1)

Table 17. Thermodynamic parameters in the binding reaction of NLS-NUPR1 peptides with Impα3 obtained by ITC at 25 °C. The parenthesis in n parameter indicate that these values were fixed in order to obtain convergence in the curve fit due to low affinity.

Peptide	ΔImpα3				
	K_a (M^{-1})	K_d (μM)	ΔH (kcal/mol)	ΔG (kcal/mol)	n
WT	$1.1\cdot10^6$ [$0.5\cdot10^6$, $1.5\cdot10^6$]	0.95 [0.55, 1.35]	-3.7 [-4.2, -3.2]	-8.2 [-8.7, -7.7]	1.0 [0.8, 1.2]
K65A	$3.7\cdot10^5$ [$2.9\cdot10^5$, $4.5\cdot10^5$]	2.7 [2.4, 3.1]	-10.2 [-10.7, -9.7]	-7.6 [-8.1, -7.1]	1.4 [1.2, 1.6]
K69A	$1.3\cdot10^5$ [$0.8\cdot10^5$, $2.1\cdot10^5$]	7.6 [7.1, 8.1]	-21.3 [-21.8, -20.8]	-7.0 [-7.5, -6.5]	1.4 [1.2, 1.6]
T68E	$8.3\cdot10^4$ [$7.5\cdot10^4$, $9\cdot10^4$]	12.0 [11.5, 12.5]	-17.5 [-18.0, -17.0]	-6.7 [-7.2, -6.2]	(1)
K65AT68E	$7.1\cdot10^4$ [$6.7\cdot10^4$, $7.5\cdot10^4$]	14.0 [13.5, 14.5]	-17.9 [-18.4, -17.4]	-6.6 [-7.1, -6.1]	(1)
T68EK69A	$5.9\cdot10^4$ [$5.3\cdot10^4$, $6.5\cdot10^4$]	17.0 [16.5, 17.5]	-21.2 [-21.7, -20.7]	-6.5 [-7.0, -6.0]	(1)
K65AT68EK69A	$4.2\cdot10^4$ [$3.7\cdot10^4$, $4.7\cdot10^4$]	24.0 [23.5, 24.5]	-28.5 [-29.0, -28.0]	-6.3 [-6.8, -5.8]	(1)
pT68	$3.4\cdot10^4$ [$3\cdot10^4$, $3.9\cdot10^4$]	29.0 [28.5, 29.5]	-28.2 [-28.7, -27.7]	-6.2 [-6.7, -5.7]	(1)

Table 18. Thermodynamic parameters in the binding reaction of NLS-NUPR1 peptides with ΔImpα3 obtained by ITC at 25 °C. The parenthesis in n parameter indicate that these values were fixed in order to obtain convergence in the curve fit due to low affinity.

The results obtained in these ITC experiments indicated that the affinity of these mutant peptides for importins was lower comparing to the WT peptide. Concretely:

- The WT peptide exhibited the highest affinity for both Impα3 and ΔImpα3.
- In general, most of the peptides showed higher affinity for binding to ΔImpα3, except for the T68EK69A and pT68 peptides.
- The removal of either Lys65 or Lys69 residues from the peptides led to a decrease in affinity. Notably, the variations in affinity were higher for ΔImpα3 than for Impα3.
- Phosphorylation or mutation of Thr68 to Glu (phospho-mimics) resulted in a significant decrease in affinity, approximately one order of magnitude lower, compared to the other mutations for both importin species.

4.1.5. NUPR1L interacted with Impα3 and ΔImpα3 *in vitro*

As NUPR1, its paralogue NUPR1L must be translocated from the cytoplasm to the nucleus to carry out its functions. Similar to NUPR1, it was thought that Impα3 might be responsible for internalizing NUPR1L into the nucleus, so it was intended to follow the same approach as with NUPR1 and evaluate the affinity of NUPR1L for Impα3 and ΔImpα3 by ITC. However, as already observed in previous studies, NUPR1L was highly prone to aggregation (319). This tendency to aggregation could be the cause of producing thermograms exhibiting a significant peak upon NUPR1L dilution, which hampered the possibility of assessing the binding to both Impα3 and ΔImpα3 by ITC.

Fortunately, the affinity of NUPR1L for both importins was measured by fluorescence thanks to the collaborations mentioned at the beginning of Chapter II. Fluorescence results (data not shown, only K_d values are detailed. For further information: (336)) indicated that NUPR1L interacted with both importin species, with apparently similar K_d values in the low micromolar range (K_d of 4.0 ± 0.7 µM for Impα3, and K_d of 5.0 ± 1.0 µM for ΔImpα3). Therefore, these results verified that the NUPR1 paralogue was also able to interact with Impα3 with a high affinity, indicating that the internalization of NUPR1L to the cell nucleus also must occur by the mechanism of the classical nuclear protein import carried out by importins.

4.1.6 The NLS-NUPR1L peptide interacted with both Impα3 and ΔImpα3 *in vitro*

The following step was to analyze whether the theoretically predicted NLS region of NUPR1L was capable of binding in isolation to both Impα3 and ΔImpα3. Therefore, it was evaluated the affinity of NLS-NUPR1L peptide for both importins by ITC (**Figure 81 and Table 19**).

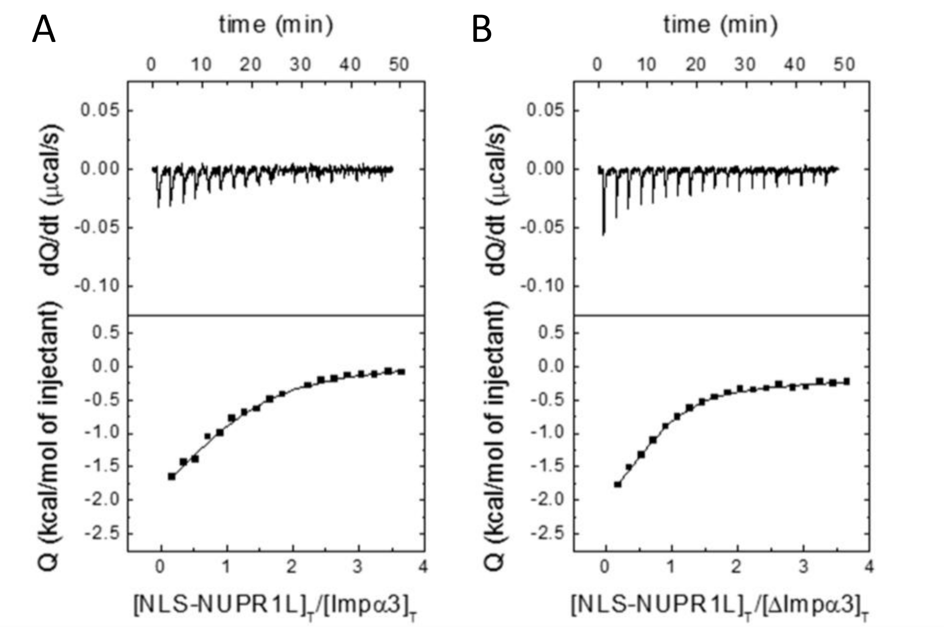


Figure 81. The interaction NLS-NUPR1L peptide with (A) Impα3 and (B) ΔImpα3 was measured by ITC. Thermograms (upper panels; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherms (lower panels; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fits according to a model considering a single ligand binding site (continuous line) are shown.

	K_a (M^{-1})	K_d (μM)	ΔH (kcal/mol)	ΔG (kcal/mol)	n
Impα3/NLS-NUPR1L	$8.3\cdot10^{4}$ $[7.9\cdot10^{4}, 8.7\cdot10^{4}]$	12.0 [10, 14]	-3.1 [-3.6, -2.6]	-6.7 [-7.2, -6.2]	1.04 [0.99, 1.09]
ΔImpα3/NLS-NUPR1L	$1.8\cdot10^{5}$ $[1.3\cdot10^{5}, 2.1\cdot10^{5}]$	5.5 [4.6, 6.4]	-2.4 [-2.9, -1.9]	-7.2 [-7.7, -6.7]	0.75 [0.69, 0.81]

Table 19. Thermodynamic parameters in the binding reaction of NLS-NUPR1L with Impα3 and ΔImpα3 obtained by ITC at 25 °C.

These results indicated that the NLS-NUPR1L isolated region was able to interact with both importin species. Specifically, the affinity of NLS-NUPR1L for ΔImpα3 (K_d of 5.5 μM) was higher than that for Impα3 (K_d of 12.0 μM). Furthermore, due to the fact that it bound with similar affinities than intact NUPR1L, these results confirmed, first, that NUPR1L contained a NLS region in its sequence; secondly, that the key residues responsible for the binding between NUPR1L and both Impα3 and ΔImpα3 were located in this sequence patch and; third, that the IBB domain removal favored the binding of NUPR1L to Impα3.

Therefore, these results allowed an indirectly characterization of the binding between NUPR1L and both Impα3 and ΔImpα3, thanks to the study of the binding parameters between the NLS sequence of NUPR1L and both importins.

4.1.7. ZZW-115 inhibited the nuclear transport of NUPR1 by competing with importins

As previously explained in the introduction to Chapter II, the TFP-derived compound ZZW-115 showed high affinity for NUPR1 *in vitro*, exhibiting a 10-fold higher antitumor activity than TFP. Furthermore, in assays with xenograft mice, a dose-dependent tumor regression caused by ZWW-115 compound was observed, even causing tumors to disappear after 30 days of treatment (331). During the present Doctoral Thesis, knowing already that ZZW-115 had no apparent neurological effects, but not yet knowing that it bound to the hERG channel and could cause cardiotoxicity in humans, it was considered crucial to determine the molecular mechanism by which the ZZW-115 compound exerted its antitumor activity.

Therefore, once verified that the NLS sequence of NUPR1 was essential for its binding to Impα3 and nuclear translocation, and due the fact that, first, in previous studies of our group it was determined that NUPR1 bound to the ZZW-115 compound by using the residues around Ala33 and Thr68 (the two "hot spots" of NUPR1) (331), and second, that Thr68 belongs to the NLS region of NUPR1, we evaluated whether ZZW-115 was able to prevent the interaction between NUPR1 and Impα3. Thus, the compound ZZW-115 could block the nuclear translocation of NUPR1 through its NLS sequence, thereby preventing it from carrying out its biological functions. Consequently, possible differences between the NUPR1/Impα3 interaction in presence of the ZZW-115 compound, comparing to this interaction in absence of ZZW-115, were assessed by ITC (**Figure 82**).

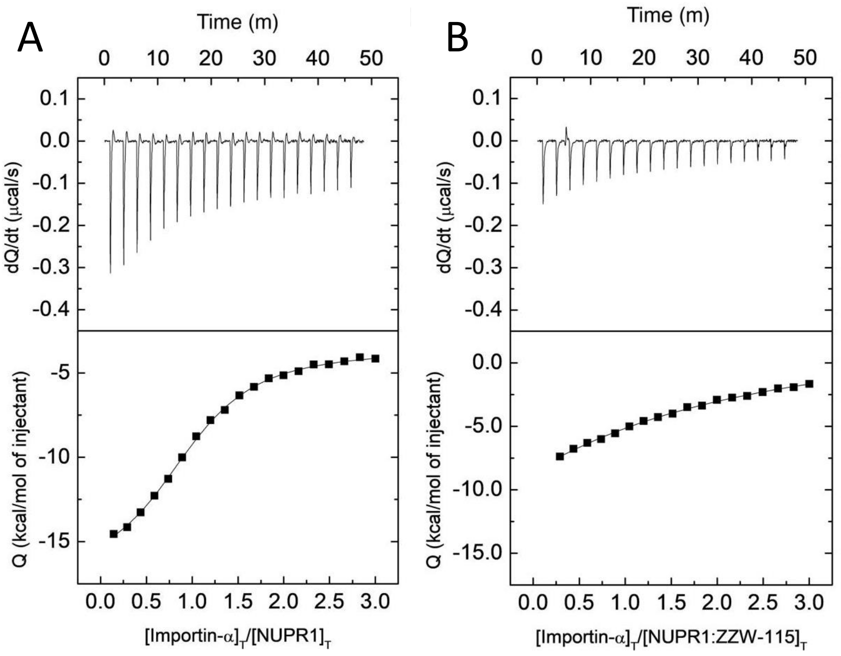


Figure 82. ITC raw data for the interaction between Impα3 and NUPR1 in the absence (A) or presence (B) of ZZW-115. Thermograms (upper panels; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherms (lower panels; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fits according to a model considering a single ligand binding site (continuous line) are shown.

In the ternary assay, a 100 µM concentration of ZZW-115 produced a 25-fold reduction in the affinity between NUPR1 and Impα3 (**Table 20**).

Impα3/NUPR1	K_a (M^{-1})	K_d (µM)	ΔH (kcal/mol)	ΔG (kcal/mol)
- ZZW-115	$6.9 \cdot 10^5$ [$6.2 \cdot 10^5$, $7.6 \cdot 10^5$]	1.4 [0.9, 1.9]	-13.7 [-14.2, -13.2]	-8.0 [-8.5, -7.5]
+ ZZW-115	$2.8 \cdot 10^4$ [$2.3 \cdot 10^4$, $3.3 \cdot 10^4$]	35.0 [34.5, 35.5]	-61 [-61.5, -60.5]	-6.1 [-6.6, -5.6]

Table 20. Thermodynamic parameters for the Impα3/NUPR1 interaction, in the absence or presence of ZZW-115, determined by ITC at 25°C.

Therefore, together with other results that demonstrated that ZZW-115 inhibited NUPR1 activity through its interaction with Thr68 of the NLS sequence of NUPR1 (292), it was verified that the affinity of Impα3 for NUPR1 was significantly reduced in the presence of ZZW-115 by ITC, thereby hampering the NLS recognition by Impα3 and nuclear translocation of NUPR1.

4.2. Identification of novel antineoplastic compounds targeting the NUPR1 protein

4.2.1. Experimental screening procedure to identify potential ligands targeting NUPR1

As previously mentioned, several studies of CD and NMR spectra showed that NUPR1 is mostly unfolded (298,337) Despite that, there is also evidence of a local, labile structure, that might be stabilized by interacting ligands (338).

When a ligand interacts with a protein, this interaction might induce some limited structural rearrangements, which could result in different thermal denaturation patterns comparing to the unliganded protein. These changes can be monitored by fluorescence using an extrinsic probe such as ANS. In this way, it was possible to employ ligand-induced stabilization against thermal denaturation to identify potential NUPR1 interacting compounds. The ANS extrinsic probe binds to hydrophobic patches or hydrophobic residues of proteins (214), and when this occurs, an increase in the fluorescence intensity can be generally recorded upon protein unfolding. However, depending on the hydrophobicity changes produced in the surface area exposed to the solvent of a protein or protein complexes, sometimes a decrease in the fluorescence intensity upon protein binding can be recorded.

A collection of 10,000 compounds (Maybridge Chemical Library) was screened *in vitro*, following a procedure based on thermal denaturations of NUPR1 protein, using TSA. From this HTS, 45 compounds were identified and selected, from now on named AJO1 to AJO45, as those able to produce different temperature denaturation profiles in NUPR1 (**Figure 83**), when compared to controls (NUPR1 without compound) (**Figure 84**).

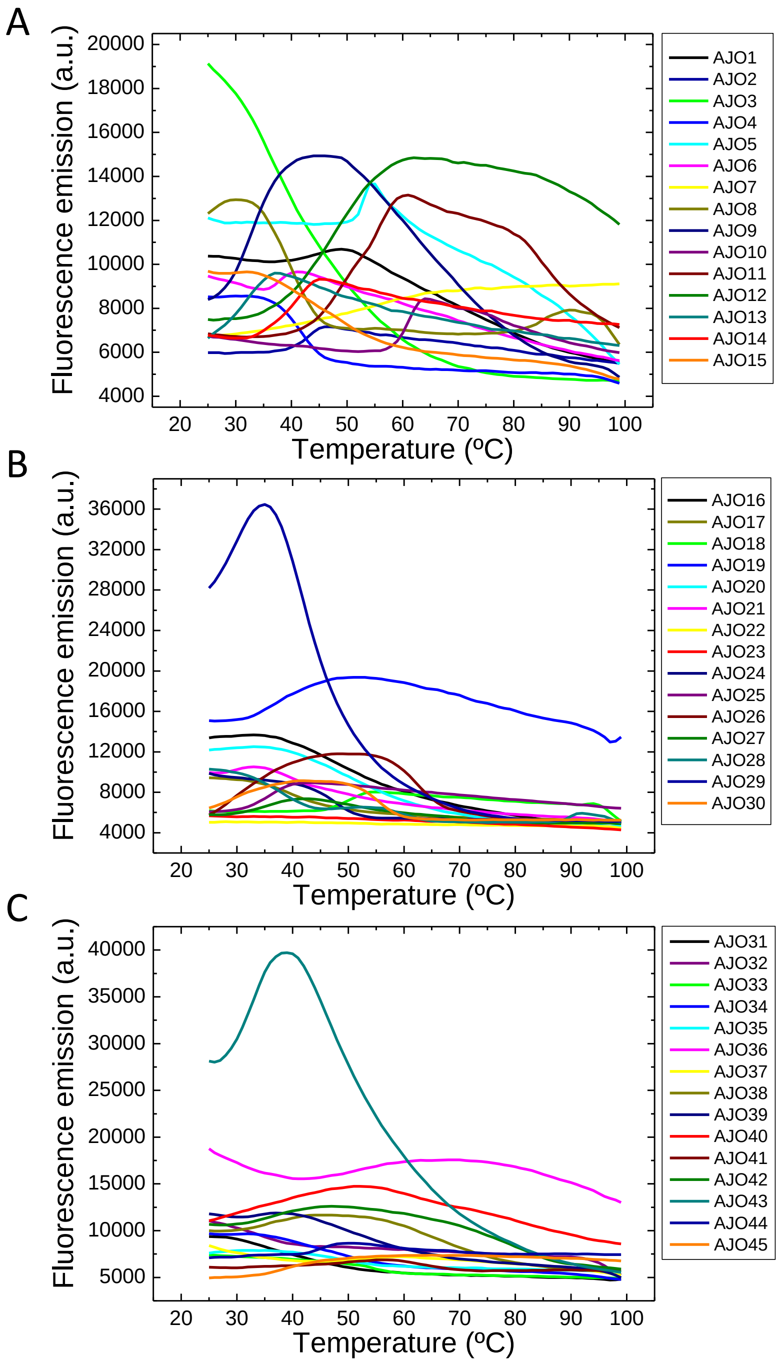


Figure 83. *In vitro* molecular HTS of compounds based on thermal denaturations of NUPR1 by TSA. Selection of compounds capable of interacting with NUPR1 and altering the NUPR1 thermal denaturation profile. For a better visualization of the curves, the graphs were divided into three groups, as follows: (A) AJO1-AJO15 compounds, (B) AJO16-AJO30 compounds, (C) AJO31-AJO45 compounds.

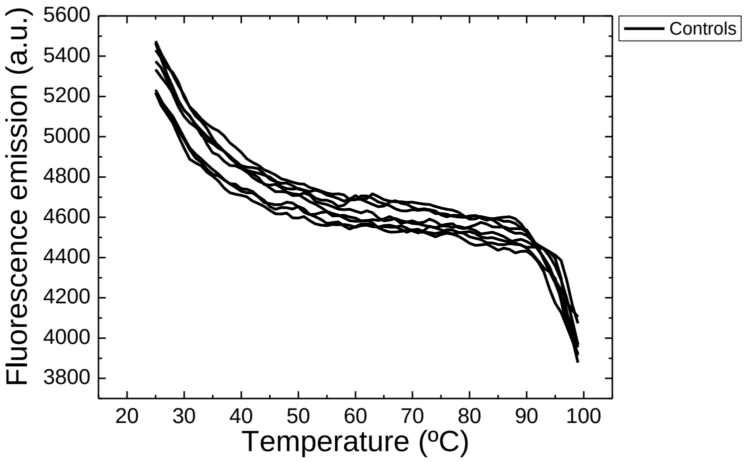


Figure 84. Controls, i.e., NUPR1 without compounds but with DMSO, obtained in the *in vitro* molecular screening of compounds based on thermal denaturations of NUPR1 by TSA. Of note, the scale in the Y-axis is very different to that in the TSA plots with compounds (Figure 83).

Unlike in Chapter I of this Doctoral Thesis, the specific names of the selected compounds, nor of their therapeutic properties/indications are not detailed in Chapter II, because of confidentiality reasons and also due to the fact that they were compounds selected from the Maybridge chemical library, which did not contain FDA-approved drugs with known therapeutic indication.

4.2.2. Target engagement of selected compounds by ITC

After identifying 45 compounds through the *in vitro* experimental HTS based on ligand-induced stabilization of NUPR1, it was crucial to confirm their interaction with NUPR1 and determine their binding affinity. Of note, the protein structural stabilization effect does not always directly correlate with ligand binding affinity, and compounds that induce a higher stabilization effect may not necessarily exhibit higher affinity. Additionally, protein stability can increase (or apparently increase) due to non-specific interactions between the compound and the protein, or intrinsic properties of the compound (such as fluorescence emission, color, temperature-dependent aggregation, etc.).

To directly measure the binding affinity of NUPR1 for the selected compounds, ITC assays were performed. While ITC assays were performed for all the AJOX compounds, only the calorimetric titration of AJO14, the most promising compound, is presented as a representation (**Figure 85**). Of note, AJO5, AJO17 and AJO22 were excluded from further studies due to their low solubility and precipitation/aggregation-related problems, which would hinder their tractability for subsequent assays and optimization efforts. Therefore, from this moment on, 42 AJO were the remaining compounds for the following assays.

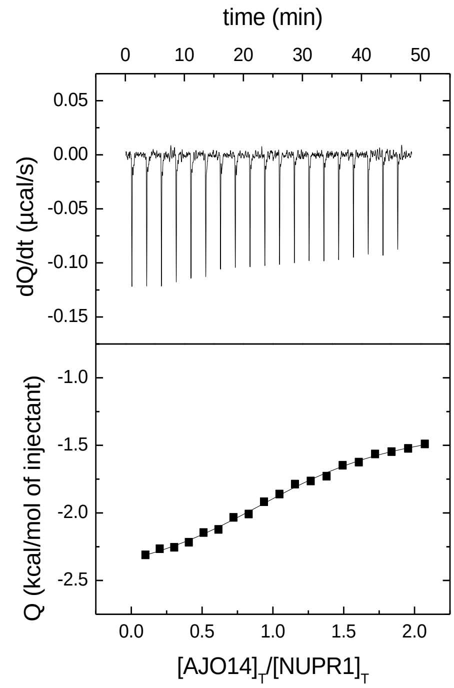


Figure 85. Calorimetric titration for AJO14 interacting with NUPR1. Thermogram (upper panel; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherm (lower panel; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fit according to a model considering a single ligand binding site (continuous line) is shown.

The ITC results revealed that AJO14 exhibited a K_d in the low micromolar range, indicating a favorable binding affinity (**Table 21**). This finding suggested that AJO14 could serve as a suitable starting point for further optimization in terms of structural and functional properties.

Compound	**K_a (M^{-1})**	**K_d (µM)**	**ΔH (kcal/mol)**	**ΔG (kcal/mol)**	**n**
AJO14	$2\cdot10^5$ [$1.3\cdot10^5$, $3\cdot10^5$]	4.9 [3.3, 7.8]	-1.2 [-1.5, -1.1]	-7.2 [-7.5, -6.9]	1.2 [1.1, 1.3]

Table 21. Thermodynamic parameters of the NUPR1-AJO14 complex obtained by ITC at 15 °C.

4.3. Effects of the selected compounds on the NUPR1 protein using *in vitro* and *in vivo* systems

4.3.1. Selected compounds induced a growth suppressive effect in different cell lines

In the evaluation of the anticancer activity, the 42 compounds selected in the *in vitro* experimental screening were subjected to the PrestoBlue^{TM} cell viability assay to evaluate changes in the cell proliferation rate. For this purpose, MiaPaCa-2 cells were treated with several concentrations of AJO compounds, ranged from 0-100 µM, and the ZZW-115 compound as reference of a compound with a great inhibition of cell proliferation capacity. Among all these compounds, AJO14, AJO23, and AJO40 demonstrated notable antiproliferative activities due to the fact that they reduced the viability of MiaPaCa-2 cells (**Figure 86**), also reflected in their IC_{50} values (**Table 22**). Specifically, AJO40 exhibited a promising low IC_{50} value, being only slightly higher than ZZW-115. However, it should be noted that the overall anticancer activities of these compounds were not as potent as the previously discovered NUPR1 inhibitor ZZW-115 (331).

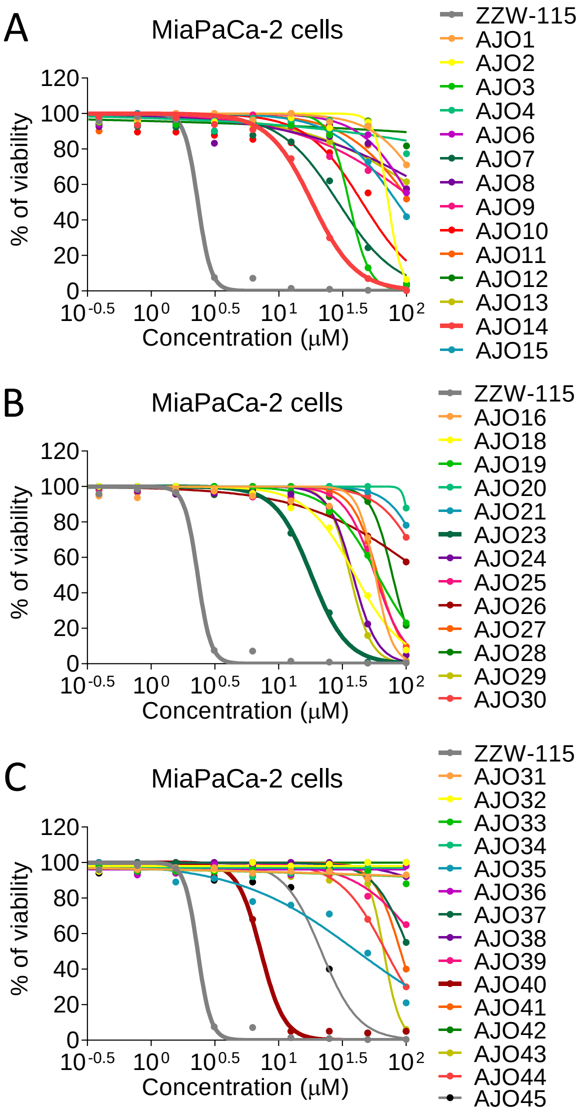


Figure 86. Quantification of the *in vitro* inhibitory effects of the 42 selected top compounds, and ZZW-115 as reference compound, on proliferation of MiaPaCa-2 cells measured by PrestoBlue™ cell viability assay. For a better visualization of the results, ZZW-115, AJO14, AJO23 and AJO40 compounds are in bold type, and the graphs were divided into three groups, as follows: (A) AJO1-AJO15 compounds, (B) AJO16-AJO30 compounds, (C) AJO31-AJO45 compounds.

Compound	IC_{50} (µM)
ZZW-115	2.3
AJO1	>100
AJO2	72.0
AJO3	35.6
AJO4	>100
AJO6	>100
AJO7	29.1
AJO8	>100
AJO9	>100
AJO10	45.0
AJO11	>100
AJO12	>100
AJO13	>100
AJO14	16.0
AJO15	84.4
AJO16	57.3
AJO18	40.0
AJO19	60.0
AJO20	>100
AJO21	>100
AJO23	18.1
AJO24	38.3
AJO25	57.2
AJO26	>100
AJO27	60.0
AJO28	78.2
AJO29	36.3
AJO30	>100
AJO31	>100
AJO32	>100
AJO33	>100
AJO34	>100
AJO35	41.2
AJO36	>100
AJO37	>100
AJO38	>100
AJO39	>100
AJO40	7.3
AJO41	90.6
AJO42	>100
AJO43	66.9
AJO44	69.5
AJO45	21.6

Table 22. IC_{50} values of the 42 AJO compounds and ZZW-115, as a reference compound, on MiaPaCa-2 cells.

To further validate the hypothesis that these compounds exerted their anticancer effects by specifically targeting NUPR1, their cytotoxicities were assessed on Panc-1 cell line clones, both NUPR1-WT and NUPR1-KO clones (**Figure 87**), which were developed using the CRISPR-Cas9 technology, and their IC_{50} values were calculated (**Table 23**).

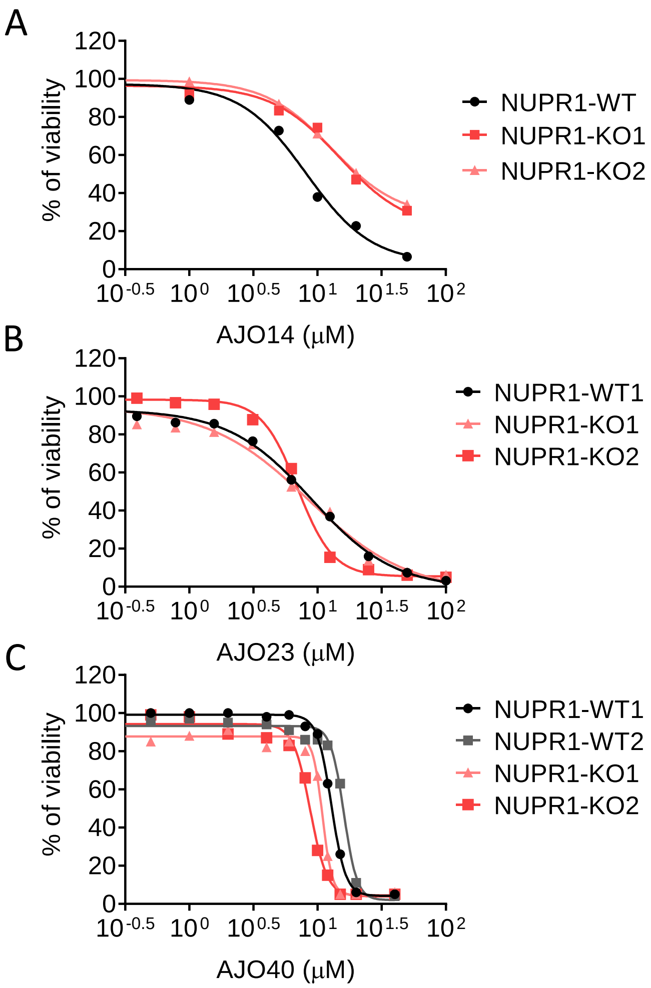


Figure 87. Viability percentage of NUPR1-WT and NUPR1-KO clones of Panc-1 cells treated with (A) AJO14, (B) AJO23 or (C) AJO40 for 72 h. Experiment were performed in duplicate.

Compound	**IC_{50} (µM)**			
	NUPR1-WT1 clone	**NUPR1-WT2 clone**	**NUPR1-KO1 clone**	**NUPR1-KO2 clone**
AJO14	8.2	-	14.9	13.0
AJO23	8.9	-	8.4	7.0
AJO40	12.8	16.0	10.9	8.8

Table 23. IC_{50} values of the three top AJO compounds on Panc-1 cell line clones.

The results revealed that the anticancer activity of AJO14 on NUPR1-WT cells was 2-3 times higher compared to NUPR1-KO cells. Conversely, AJO23 exhibited similar antiproliferative activity on both NUPR1-WT and NUPR1-KO cells, therefore indicating that AJO23 was not specific for NUPR1 and it should have another mechanism of action for reducing cell proliferation. Unfortunately, despite the fact that AJO40 was the most effective compound in suppressing cell growth on MiaPaCa-2 cells (IC_{50} = 7.3 µM), in the assay with Panc-1 clones, it was observed that

KO clones had lower viability than WT clones, therefore AJO40 was not specific for NUPR1. Then, despite being a highly potent compound, as it did not show specificity against the target under study, it had to be discarded for further assays (albeit, it could be interesting to pursue AJO40 in other directions). Based on these findings, it can be inferred that AJO14 was the most likely compound to exert its anticancer effects by specifically binding to NUPR1. Therefore, AJO14 was the compound chosen after these techniques for further assays.

Further evaluation of the antitumor activity of the AJO14 was tested on 11 different primary PDAC-derived cell lines, revealing its potent effectiveness against PDAC cells (**Figure 88**) with IC_{50} values ranging from 10.0 µM to 22.2 µM (**Table 24**).

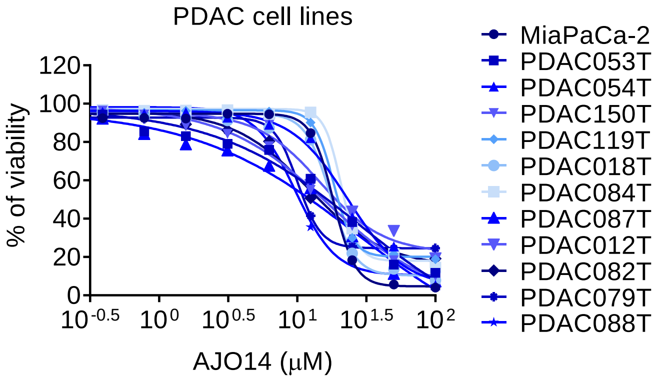


Figure 88. Viability of 11 primary PDAC-derived cell lines treated with AJO14 for 72 h comparing to MiaPaCa-2 cells viability.

Primary PDAC-derived cell line	IC_{50} (µM)
MiaPaCa-2	16.0
PDAC053T	18.5
PDAC054T	22.2
PDAC150T	16.8
PDAC119T	18.4
PDAC018T	17.3
PDAC084T	20.8
PDAC087T	18.2
PDAC012T	19.5
PDAC082T	12.1
PDAC079T	10.0
PDAC088T	10.5

Table 24. IC_{50} values of AJO14 on 11 primary PDAC-derived cell lines.

Due to the fact that NUPR1 is highly expressed in several different cancer cells, the antitumor activity of AJO14 was also evaluated on other cancer cell lines to evaluate its antitumor effect on other types of tumors. These different cancer cell lines included HepG2 (hepatocarcinoma), MDA-MB-231 (breast cancer), H358 (lung cancer), A375 (melanoma), HT-29 (colon cancer), U87 (glioblastoma), U2OS (osteosarcoma), and PC-3 (prostate cancer). The results indicated that AJO14 effectively induced cell death in these diverse cancer cell types (**Figure 89**), with IC_{50} values ranging from 10.3 µM to 22.0 µM (**Table 25**), therefore not being specific to a particular cell type.

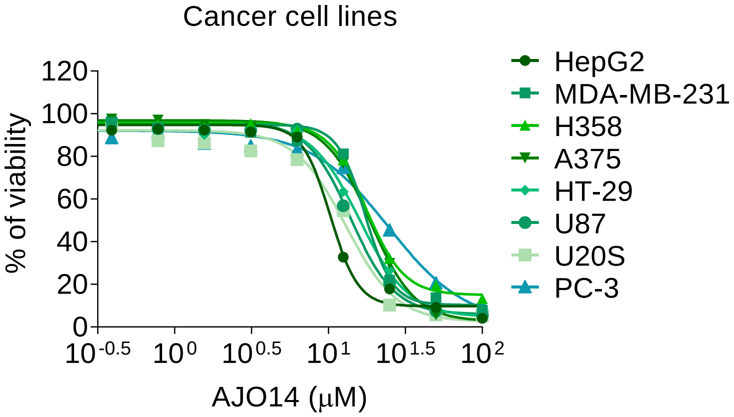


Figure 89. Viability of different cancer cell lines (HepG2, MDA-MB-231, H358, A375, HT-29, U87, U2OS and PC-3) after treating them at increasing concentrations of AJO14 for 72 h.

Cancer cell line	**IC_{50} (µM)**
HepG2	10.3
MDA-MB-231	19.0
H358	20.2
A375	18.0
HT-29	17.0
aU87	14.0
U20S	13.0
PC-3	22.0

Table 25. IC_{50} values of AJO14 on several cancer cell lines.

In summary, AJO14 demonstrated the ability to not only reduce viability of PDAC cells but also exert its anticancer effect on other cancer cell lines by specifically targeting NUPR1. Therefore, these results were promising for further *in vitro* assays with the AJO14 compound in order to gain more insight into its mechanism of action in reducing the cell viability of cancer cells.

4.3.2. AJO14 showed very low hERG binding ability

In the drug discovery process, one of the most important reasons contributing to drug attrition and failure is the appearance of cardiac arrhythmia, which is a commonly observed serious side effect. This problem is primarily associated with hERG channel inhibition exerted by the drug (334,335). To date, approximately 17 FDA-approved drugs have been withdrawn from the market due to their ability to block the hERG channel (339,340), which highlights the significant impact that hERG channel inhibition can have on drug safety and efficacy. Therefore, early identification of compounds with potential hERG inhibitory effects is essential.

An assay to evaluate the ability of AJO14 to block the hERG channel using the Predictor^{TM} hERG fluorescence polarization assay kit was performed. The AJO14 compound inhibitory effect was compared with that of E-4031, which is a positive control commonly used in this assay (**Figure 90**).

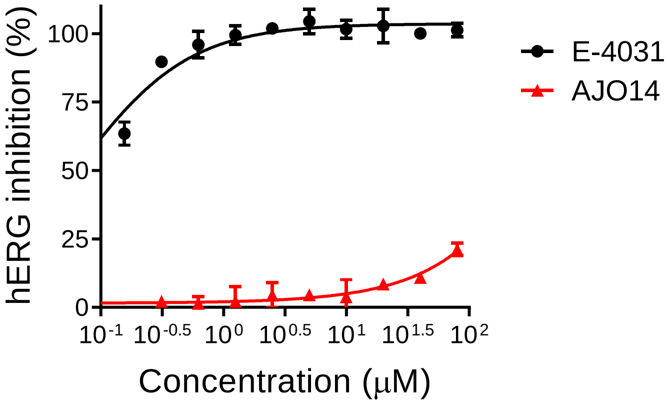


Figure 90. Concentration-response curves for hERG binding of both E-4031 and AJO14 using the fluorescence polarization assay. Data are expressed as the mean ± SD, and the experiment was performed in triplicate.

The results revealed that IC_{50} values for E-4031 and AJO14 were found to be 0.06 μM and higher than 80.0 μM, respectively, thus indicating that AJO14 was over 1,000 times less potent in inhibiting the hERG channel compared to E-4031 (**Table 26**).

Compounds	hERG channel inhibition: IC$_{50}$ (µM)	% Inhibition at 10 µM
E-4031	0.06	99%
AJO14	>80.0	4%

Table 26. IC$_{50}$ values for the inhibition of the hERG channel by E-4031 and AJO14.

Based on these results, it was concluded that AJO14 was unlikely to cause cardiotoxicity related to hERG binding, allowing this compound to continue through the development pipeline.

4.3.3. AJO14 induced PC cell death by concomitant necrosis and apoptosis

Once the efficacy of AJO14 in reducing the cell growth of different cancer cells was proven, several techniques were performed to elucidate the molecular mechanisms by which this antitumor activity was exerted. For this purpose, it was evaluated whether AJO14 induced necrosis and apoptosis in MiaPaCa-2 cells after 24, 48 and 72 h of treatment, by evaluating LDH release (CytoTox-ONE^{TM} homogeneous membrane integrity assay) and caspase-3/7 activity (Caspase-Glo® 3/7 assay), respectively.

The results showed a significant increase in the LDH release from MiaPaCa-2 cells treated with AJO14 in a concentration-dependent manner compared to the control group (**Figure 91**). Similarly, the caspase-3/7 activity in MiaPaCa-2 cells treated with AJO14 also remained higher than that in the control group (**Figure 92**).

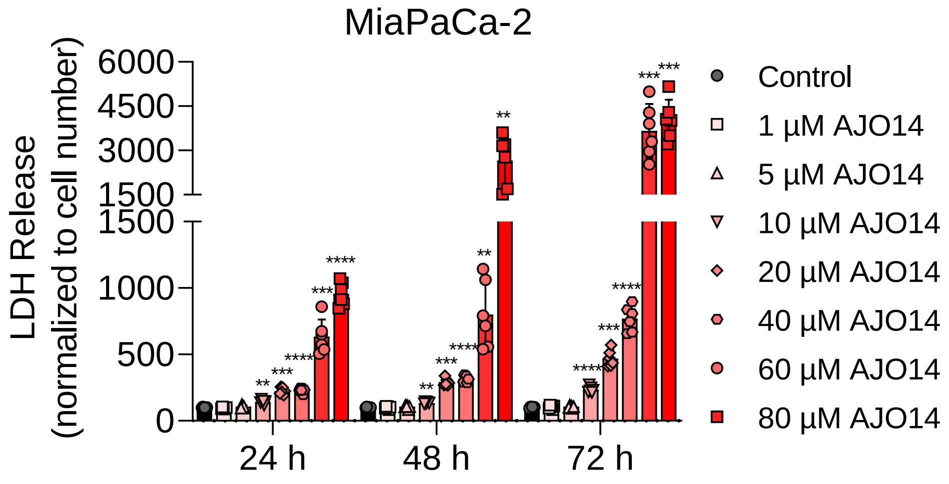


Figure 91. LDH release measurement in MiaPaCa-2 cells treated with increasing concentrations of AJO14 for 24, 48 and 72 h. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ (one-way ANOVA, Dunnett's multiple comparisons test).

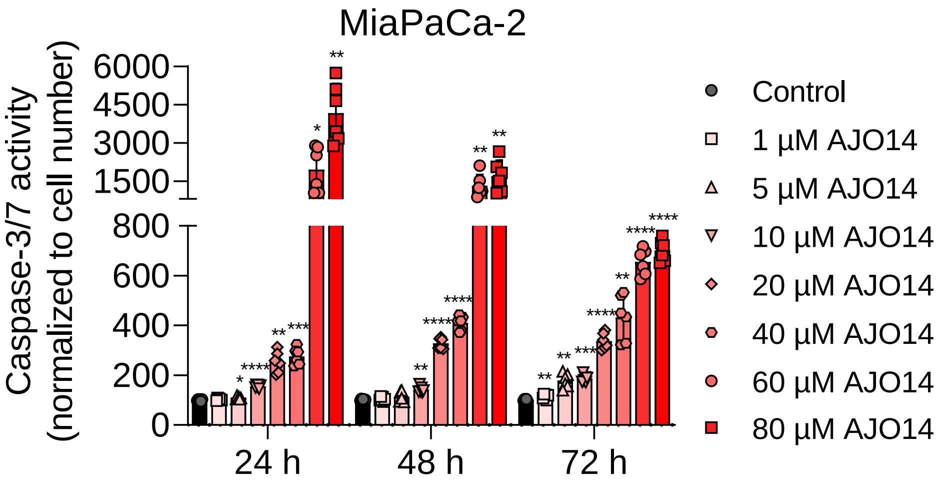


Figure 92. Caspase-3/7 activity measurement in MiaPaCa-2 cells treated with increasing concentrations of AJO14 for 24, 48 and 72 h. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ (one-way ANOVA, Dunnett's multiple comparisons test).

Since an increase in LDH release is associated with necrosis-mediated cell death, and increased caspase-3/7 activity in cells is related to apoptosis-mediated cell death, these results collectively indicated that the AJO14 compound induced programmed cell death exerting pro-necrotic and apoptotic effects.

4.3.4. AJO14 treatment induced energetic metabolic failure

Intracellular ATP production within the mitochondria is essential for an appropriate execution of the cellular functions and it also plays a crucial role in cell death mechanisms (341). Therefore, another possible effect underlying AJO14 activity could be associated to changes at the level of ATP production and release.

To investigate this, the intracellular ATP content in MiaPaCa-2 cells after 24, 48 and 72 h of treatment with increasing concentrations of AJO14 was measured (CellTiter-Glo® Luminescent Cell Viability Assay). Remarkably, a significant concentration-dependent decrease in ATP levels was observed (**Figure 93**).

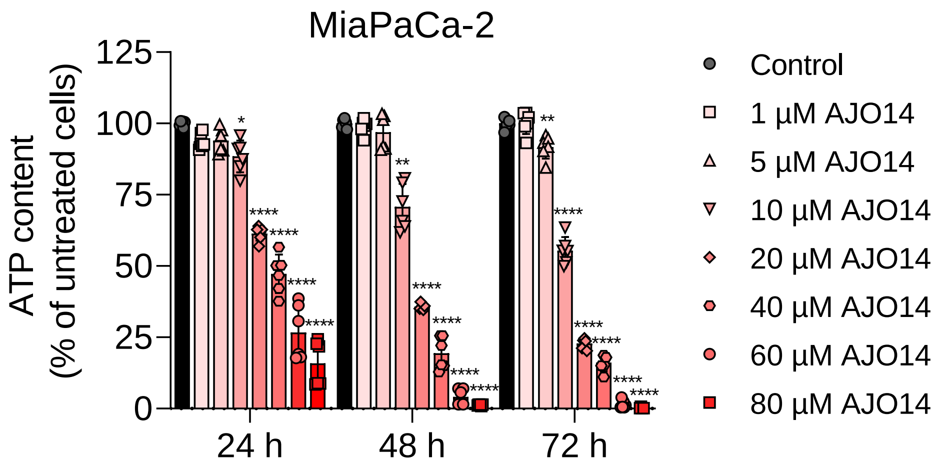


Figure 93. ATP content measurement in MiaPaCa-2 cells treated with increasing concentrations of AJO14 for 24, 48 and 72 h. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ (one-way ANOVA, Dunnett's multiple comparisons test).

Therefore, these results demonstrated that AJO14 treatment induced cell metabolic failure characterized by a decrease in ATP production.

4.3.5. AJO14 inhibited PDAC tumor development in animal models

One of the key steps in making the leap from *in vitro* to *in vivo* studies is to have performed numerous positive *in vitro* assays in order to correctly select the appropriate compound/s to be evaluated in *in vivo* animal models. Because AJO14 was the most promising candidate for targeting cancer cells *in vitro*, its *in vivo* anticancer efficacy using a MiaPaCa-2 cell xenograft mouse model was evaluated following previously described methods in recent studies of our group (331).

More specifically, after the s.c. tumor inoculation with MiaPaCa-2 cells in mice, once the tumors reached 200 mm^3, they were treated i.p. with AJO14 at different doses (5, 10, 20 and 50 mg/kg) for 28 days. For an appropriate comparison, the control group received an equivalent volume of sunflower seed oil and the same percentage of DMSO than the other groups. During the 28 days of the assay, tumor volumes in the control group displayed exponential growth, from 222.0 ± 18.2 mm^3 to 1,589.2 ± 183.1 mm^3. In contrast, mice injected with 50 mg/kg AJO14 compound exhibited an important deceleration in tumor growth, from 246.4 ± 47.7 mm^3 to 225.0 ± 53.6 mm^3. Furthermore, the other three groups treated with 5 mg/kg, 10 mg/kg, and 20 mg/kg AJO14 showed intermediate tumor growth between the control group and the 50 mg/kg AJO14 group. In particular, tumor sizes increased from 269.7 ± 107.4 mm^3 to 1,130.0 ± 389.8 mm^3 in mice treated with 5 mg/kg AJO14, from 222.5 ± 37.1 mm^3 to 933.3 ± 199.9 mm^3 in mice treated with

10 mg/kg AJO14, and from 230.8 ± 81.9 mm^3 to 858.2 ± 196.9 mm^3 in mice treated with 20 mg/kg AJO14 (**Figure 94**).

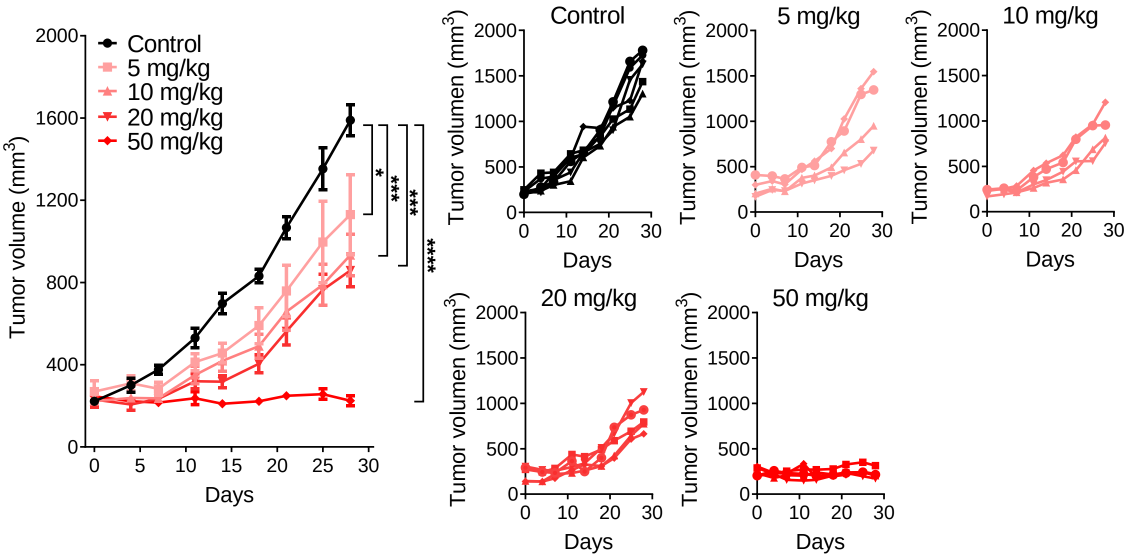


Figure 94. PDAC MiaPaCa-2 xenograft mice were i.p. daily treated with 5, 10, 20 and 50 mg/kg AJO14, and mice in the control group were treated with sunflower seed oil (5 mice were employed in each group). Tumor volumes of mice were measured twice per week. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

These findings definitively showed the potent ability of AJO14 to suppress tumor growth in a dose-dependent manner. Furthermore, it is worth noting that the administration of AJO14 to mice, regardless of the amount used, did not result in any decrease in body weight (**Figure 95**). Thus, body weights were from 21.35 ± 0.92 g to 27.32 ± 0.58 g in control mice, from 22.25 ± 0.66 g to 27.13 ± 1.49 g in mice treated with 5 mg/kg AJO14, from 20.83 ± 1.39 g to 25.90 ± 1.04 g in mice treated with 10 mg/kg AJO14, from 19.84 ± 2.39 g to 25.18 ± 3.18 g in mice treated with 20 mg/kg AJO14, and from 20.42 ± 1.75 g to 27.28 ± 2.16 g in mice treated with 50 mg/kg AJO14.

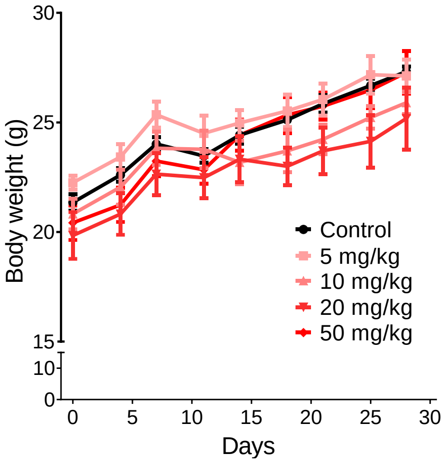


Figure 95. PDAC MiaPaCa-2 xenograft mice were i.p. daily treated with 5, 10, 20 and 50 mg/kg AJO14, and mice in the control group were treated with sunflower seed oil (5 mice were employed in each group). Body weights of mice were measured twice per week.

Finally, another really important aspect was that there were no indications of systemic toxicity, after sacrificing the mice and inspecting a large number of organs, when AJO14 was administered in mice. Altogether, these findings collectively demonstrated the capacity of AJO14 to effectively regress tumor growth in mice without inducing any negative side effect.

5. DISCUSSION

PC is one of the most aggressive and fatal cancers, characterized by distressingly poor therapeutic outcomes (11). For years, the fight against this disease has been one of the main aims of numerous research groups, which is mainly based on gaining a better understanding of the key associated molecular mechanisms, as well as on the search for new and more effective secure therapies than the current ones. Therefore, the main purposes are to achieve an improvement in the quality of life of affected patients, more potent specific treatments that improve the current survival rate, and a better diagnosis at earlier stages of the disease. To this end, numerous multidisciplinary research projects have been and continue to be carried out. One of these areas is based on identifying possible therapeutic targets involved in the development of PC, with the aim of blocking functions, thus reducing the wide variety of effects produced in this pathology. Among all the therapeutic targets, protein targets play a very important role and are the main focus of attention in many research studies. Besides, the interest in searching for drugs against a special type of protein targets, the IPDs, is increasing exponentially due to they are associated to key cellular processes (274).

In this context, due to the fact that one of the important aspects occurring in PC is that it is a disease characterized by high levels of stress (262–264), and since *nupr1* is a stress-induced gene (287), the stress protein NUPR1 seems a promising therapeutic protein target to treat PDAC. NUPR1 was initially characterized as being activated in the exocrine pancreas during pancreatitis (342), which is an inflammatory disorder that, in its chronic manifestation, behaves as a preneoplastic state for PDAC formation. Furthermore, it has been observed that NUPR1 is overexpressed under stress conditions caused by many stimuli, not only in PC, but also in a large variety of different human cancers and in most cell types, where its persistent expression plays a crucial role in their development and progression (270,289–293). For all these reasons, our research group, in collaboration with the aforementioned groups, has been studying for years the role of NUPR1 protein in the development of PDAC, as well as searching for and identifying specific inhibitors of this protein (180,282,292,331,343,344).

Thus, considering the potential of NUPR1 as a promising therapeutic target, the initial steps of our group some years ago was to validate its crucial role in PDAC cells. In this context, previous results from our collaborators showed that NUPR1 was involved in the development of PanINs (294) and the establishment of PDAC (299) in mice. In addition, a genetic inhibition approach using siRNA on xenograft mice was employed as a proof of concept to evaluate the role of NUPR1. Notably, while tumors treated with a siRNA control exhibited continuous growth, those treated with siRNA targeting NUPR1 stopped their progression (345). Other research groups similarly demonstrated in animal models that the inactivation of NUPR1, through siRNA against

NUPR1, in hepatocellular carcinoma (304,346), cholangiocarcinoma (347), lung adenocarcinoma (348,349), osteosarcoma (350), multiple myeloma (351,352), glioblastoma (353) and ovarian cancer (354) resulted in cell death and the inhibition of tumor growth. Altogether, these findings suggested a crucial role of NUPR1 overexpression on the tumor biology, as well as NUPR1 being a promising therapeutic target not only for PDAC, but also for other type of cancers, i.e., in general cancer therapy.

Therefore, the more knowledge there is about the mechanisms by which NUPR1 carries out its functions, the more knowledge there will be to use it as a therapeutic target and to develop inhibitory compounds in the most efficient way possible. For all these objectives, the most important initial aspect is to have an optimal expression and purification system for the protein, in order to obtain it in a pure state and in large quantities. Fortunately, NUPR1 is a protein that can be expressed in *E. coli* and purified by IMAC system in a very simple way, thus obtaining high amounts of the protein in a pure state to perform the assays. Once expressed and purified, and after verifying the correct identity of this protein by MALDI-TOF/TOF, the different assays proposed were able to be carried out.

As previously mentioned, numerous studies have shown that, in order to carry out its functions, NUPR1 interacts with a large number of biological partners (180,297,298). Of note, in recent studies of our group the interactome of NUPR1 was determined (292). Among all the identified proteins, more than 30 were found to be components of the nuclear pore, including some importins, such as Impα. In eukaryotic cells there are multiple carrier proteins which roles rely on intervening in the protein transport from the cytoplasm to the nucleus (321,322), and this molecular trafficking is a crucial cellular process for cell survival and development. In this context, Impα is one of the most important carrier proteins for nuclear translocation (323). Therefore, since NUPR1 exerts its functions in the nucleus (284,288), and Impα belonged to NUPR1 interactome, it was evaluated whether NUPR1 translocated to the cell nucleus through the classical molecular mechanism of nuclear protein import (321) to play its critical role for cancer cells. For this purpose, ITC assays were performed to analyze the *in vitro* interaction between NUPR1 and Impα3, one of the most flexible Impα isoforms. This flexibility capacity provides it a greater capacity of interacting with different cargos, because of its facility to adopt a higher variety of conformations (327). Furthermore, Impα3 was easily expressed and purified, compared to other isoforms.

The affinity of Impα3 for NUPR1 was in the micromolar range in the ITC assays, with a K_d value of 1.4 μM, thus confirming the interactome result, which indicated that NUPR1 interact with importin species, and, furthermore, with a high affinity. These results were also confirmed in assays carried out by our collaborators. For instance, fluorescence assays were performed (data not shown, (292)), in which changes in the spectrum after excitation at either 280 or 295 nm were observed. Due to the fact that NUPR1 contains in its sequence solely 2 tyrosines (Tyr30 and Tyr36), the changes detected in the fluorescence spectrum by excitation at 295 nm were probably due to changes in the aromatic residues of the protein, and more specifically to changes in the environment of at least 1 of the 6 tryptophans that Impα3 contains. Apart from these *in vitro* assays, it was also demonstrated that NUPR1 interacted with Impα3 *in cellulo* by proximity ligation assay (PLA) (data not shown, (292)). Briefly, after transfecting MiaPaCa-2 cells with a plasmid expressing Impα3-Flag, it was observed by using fluorescent antibodies in the PLA technique that NUPR1 and Impα3 co-localized in these cells. Therefore, these findings were in agreement with the results obtained from both proteomic and biophysical techniques.

Once it was proven that NUPR1 interacted with Impα3, different assays were carried out to analyze which were the important residues that made this interaction possible.

The modulation of complex assembly between importins and their cargos has been attributed to the IBB domain of Impα (324). In fact, it is already known that the IBB domain has an autoinhibitory effect and hampers the entrance of NLS peptides into the major NLS-binding region of Impα (324,328). Besides, this domain is even involved in the formation of homodimeric species between importins, resulting in a reduced capacity for cargo binding efficacy (355). Therefore, some assays were performed in order to verify the molecular significance of the IBB domain in the binding of cargos to importins. In particular, it was evaluated whether NUPR1 was able to interact with the truncated Impα3 in the absence of this IBB domain, i.e., to ΔImpα3. This interaction was analyzed by ITC, and it was observed that NUPR1 also interacted with a high affinity to ΔImpα3, concretely with a K_d value of 0.44 μM. These findings, in addition to verify that NUPR1 interacted with both Impα3 and ΔImpα3, confirmed the autoinhibitory role exerted by the IBB domain of Impα3, as the affinity of NUPR1 for ΔImpα3 (0.44 μM) was higher than that for Impα3 (1.4 μM). Since years, the IBB domain has a key role in the modulation of complex formation between importins and their cargo molecules, but these results indicated that this modulation does not only occur in well-folded proteins (324).

Another important aspect is that the NLS sequence of many proteins is an essential requirement to be transported to the nucleus through the classical nuclear import pathway, since this mechanism starts when Impα recognizes the NLS in the cargo (324). Due to the importance of this NLS sequence, one of the aims of this Doctoral Thesis was to evaluate whether the theoretically predicted NLS region of NUPR1 exhibited the ability to interact in isolation with Impα3 and ΔImpα3, as well as to evaluate which residues were important for the interaction between NUPR1 and both importin species. The importance of specific residues for protein structure and function can be assessed by developing mutants of the protein under study (356,357). In this approach, while the primary focus was to understand the interaction between the full-length NUPR1 and various mutants with both Impα3 and ΔImpα3, NLS-NUPR1 peptides were employed to evaluate by ITC the specific binding mechanisms, because mutations introduced in the intact NUPR1 often resulted in poor expression, or even no expression, of the corresponding mutant.

The findings demonstrated that the isolated WT NLS-NUPR1 was able to interact with both intact Impα3 and ΔImpα3, with affinity binding constants comparable to that observed with intact NUPR1 (K_d value of 1.4 μM). In particular, the WT NLS-NUPR1 peptide showed a 2-fold higher affinity for ΔImpα3 (K_d of 0.95 μM) compared to Impα3 (K_d of 1.7 μM). This affinity was also within the same range of binding affinities observed with several other natural partners of NUPR1 (297,298,319) as well as with synthetic compounds (180,296). These results indicated firstly, that the presence of the IBB domain exerted an autoinhibitory effect (thus verifying again this effect), hampering the entry of the NLS-NUPR1 peptide into the major NLS-binding site of Impα3. This fact was in accordance with findings in previous studies involving well-folded proteins (328). Secondly, although the affinities of the WT NLS-NUPR1 peptide for both Impα3 and ΔImpα3 were slightly lower compared to those for intact NUPR1, a significant amount of the interactions implicated in importin binding could be attributed to an amino acid sequence present in the WT NLS-NUPR1 peptide. This conclusion could be explained by the fact that there were similarities in K_d values: 1.4 μM (intact NUPR1) and 1.7 μM (WT NLS-NUPR1 peptide) for Impα3, and 0.44 μM (intact NUPR1) and 0.95 μM (WT NLS-NUPR1 peptide) for ΔImpα3. And, thirdly, because of the comparable affinity constants exhibited by the WT NLS-NUPR1 peptide and NUPR1 for Impα3, this peptide could be employed as a potential lead compound for designing an inhibitor targeting its nuclear translocation.

Another aspect is that, in previously *in vivo* studies of our group, a mutant of NUPR1 at K65, K69, K76 and K77 positions was distributed throughout the entire cell, whereas the WT NUPR1 protein was exclusively located into the cell nucleus (288). Furthermore, the two lysines K65 and K69 in the NUPR1 sequence are important for nuclear translocation, according to *in vivo* studies (358). In the context of this Doctoral Thesis, some NLS-NUPR1 peptide mutants including these mutations were designed in order to study the importance of these residues in the binding to Impα. It was observed that the K65A mutation resulted in a 2-fold reduction in affinity for Impα3 (K_d value of 3.9 µM), while the K69A mutation led to a 6-fold decrease in binding affinity (K_d value of 11.0 µM). Therefore, this reduction in affinity suggested that the absence of K69 had a more pronounced impact, likely indicating that this residue established a greater number of interactions with Impα. These fact was also confirmed by docking models, in which the simulations performed by our collaborators demonstrated that both lysine residues participated in hydrophobic and polar interactions, with the latter involving their amino moieties, with residues of Impα (data not shown, (282)). Therefore, the removal of the long side chains of the lysines would unfavorably affect these contacts, consequently leading to a reduction in the binding affinity. In fact, in several structural studies it has been demonstrated the significance of lysine residues, which play crucial roles in the interaction between importins and well-folded proteins, particularly through their disordered NLS regions (327,359–361). Despite the fact that this Doctoral Thesis did not introduce any novel mechanistic findings of importin function, it was the first time in which the importance of positive charges, as in the case of folded proteins, in the cargo binding was described within the context of an IDP. Then, these assays represented the first report where the significance of such residues has been examined *in vitro* for the NLS of an IDP. In addition, recent studies indicated that IDPs could not require the presence of importins for nuclear translocation, thus making all these assays more scientifically relevant (362). Thus, these results indicated that some IDPs indeed depend on importins for their translocation into the nucleus, and the governing principles of these processes appear to be similar to those observed in well-folded protein cases.

On the other hand, a comparable reduction in affinity was observed for both of the K65A and K69A NLS-NUPR1 peptides in relation to ΔImpα3 (K_d values of 2.7 µM and 7.6 µM, respectively). However, when compared to the WT NLS-NUPR1 peptide (K_d value of 0.95 for ΔImpα3), the variation was more pronounced for both mutants when interacting with ΔImpα3 compared to Impα3. Concretely, it was observed that the K65A mutation resulted in a 3-fold reduction in affinity for Impα3 (K_d value of 2.7 µM), while the K69A mutation led to an 8-fold decrease in binding affinity (K_d value of 7.6 µM). Therefore, as it was observed with intact Impα3, the decrease in affinity was more significant for the K69A peptide. These findings suggested that, while the IBB domain maintained its independent conformational characteristics within the whole of Impα3, its removal might bring about either a structural alteration in certain regions of the ARM repeats that participate in the major NLS-binding site (thereby affecting the hydrophobic contacts critical for cargo anchoring and subsequently altering its docking) or, alternatively, the removal of the IBB domain might induce broader changes in the overall protein dynamics and stability.

In addition to the importance of the two lysines (K65 and K69) of the NLS of NUPR1 for the interaction with importins, in this Doctoral Thesis it was evaluated the role that plays the Thr68. It was already known that Thr68 acts as a key residue in the binding process of NUPR1 with any partner, whether of natural or synthetic origin (180,297,298,363). In fact, along with Ala33, Thr68 is one of the “hot spots” of NUPR1 for interacting with its partners (296–298). To analyze the importance of the Thr68 as a “hot spot” region, two different approaches, using also NLS-NUPR1 mutated peptides, were employed. First, Thr68 was mutated to Glu in order to mimic phosphorylation (T68E peptide), and second, a peptide was synthesized with the phosphorylated Thr (pT68 peptide).

Among all the NLS-NUPR1 mutants employed, the pThr68 peptide displayed the lowest affinity for Impα3 or ΔImpα3, with K_d values of 27.0 μM for Impα3 and 29.0 μM for ΔImpα3. Therefore, the phosphorylation seemed to affect the binding by potentially disrupting long-range electrostatic interactions with both importins. In relation with these results, while the WT NLS-NUPR1 peptide exhibited a higher affinity for ΔImpα3 (K_d of 0.95 μM) than that for Impα3 (K_d of 1.7 μM), the introduction of the phosphate group in the pThr68 peptide led to even more larger changes, because the affinity of the pThr68 peptide for ΔImpα3 was smaller than that observed for Impα3. Thus, these data indicated additional structural changes within the main NLS-binding domain after the IBB domain removal. Of note, around 30% of eukaryotic proteins are susceptible to phosphorylation, and most of these phosphorylation sites are located within intrinsically disordered regions, owing to their susceptibility to kinases to donate a phosphate group to the substrate molecule. In fact, phosphorylation constitutes a key regulatory mechanism in some processes such as translation and transcription, among others (283).

On the other hand, the T68E peptide, which was designed to phospho-mimic the NLS-NUPR1, also exhibited a reduced affinity for both importin species (K_d values of 22.0 μM for Impα3 and 12.0 μM for ΔImpα3) when compared to the WT NLS-NUPR1 peptide. However, this decrease was not as pronounced as that observed with the pT68 peptide (K_d values of 27.0 μM for Impα3 and 29.0 μM for ΔImpα3). Therefore, this discrepancy suggested that the phospho-mimetic modification did not induce an identical effect as the phosphorylation. Concretely, the phosphorylation at Thr68 involved the replacement of the neutral hydroxyl group (OH) with a tetrahedral phosphate group (PO_4^{2-}), thus introducing two negative charges in the sequence. This alteration in the charges consequently modified the chemical, electrostatic and steric characteristics of the Thr68 environment. Then, the presence of a double-negative charge of the phosphate group produced a different chemical context from that of the Glu phospho-mimic, which only had a negative charge. It is important to note that differences in affinities between phospho-mimics and phosphorylated threonines have been observed in several well-folded proteins (364), as well as in other IDPs (365).

In relation to the K65AT68E and T68EK69A mutant peptides, their K_d values closely resembled that of the T68E peptide. Specifically, the K65AT68E peptide had a K_d value of 21.0 µM for Impα3 and 14.0 µM for ΔImpα3, and the T68EK69A peptide a K_d value of 17.0 µM for both Impα3 and ΔImpα3, compared to the T68E peptide that had K_d values of 22.0 µM for Impα3 and 12 µM for ΔImpα3. All these results together implied that, firstly, the introduction of T68E in the interaction with importins surpassed those effects produced by substitutions of individual lysines, therefore being the role of the Thr68 more relevant for the interaction. Secondly, when a lysine was removed in conjunction with a phospho-mimicked threonine, the resulting effect was not additive for double mutants. This fact could be attributed to the remaining lysine, which could establish electrostatic interactions with the glutamic. However, when all three mutations were combined, i.e., the K65AT68EK69A NLS-NUPR1 peptide, an important decrease in the affinity was obtained (K_d values of 27.0 µM for Impα3 and 24.0 µM for ΔImpα3), therefore emphasizing the notable influence of electrostatic effects between the lysines and the phospho-mimics in the binding with the two importins. Of note, previous studies in IDPs on the importance of phosphorylation of threonines had suggested that the presence of neighboring arginines can stabilize the charge of the phosphate group, therefore promoting a stabilization effect (366). In the case of NUPR1, these lysines, rather than arginines, could play this role in stabilizing the conformation.

Because of the fact that all the affinities of the NLS-NUPR1 mutant peptides for both Impα3 and ΔImpα3 decreased when the Thr68 was phosphorylated or when the phospho-mimicking modification was introduced, it is probable that the reduction in the affinity upon phosphorylation was associated with a conformational change around the Thr68, which was likely induced by the introduction of negative charges in the sequence. Currently, the biological significance *in vivo* of this phosphorylation at the Thr68 remains unanswered, therefore being a possible future study of considerable interest. Nonetheless, recent assays performed by our research group demonstrated that the substitution of Thr68 with a Gln residue impeded the assembly of multiple complexes involving NUPR1 and other proteins related to processes of SUMOylation (292). Therefore, these new findings obtained in this Doctoral Thesis corroborated and validated these previous results related to the Thr68. At this point, it is essential to highlight that the influence of phosphorylation, in conjunction with other PTMs, can exert different effects on protein conformations (367), being really important in the IDP context (368,369) because IDPs are relevant “hot spots” for PTMs (283). Among other effects, the most important are: structural alterations at local levels (370), changes in the entropy of conformational states (371), changes in interactions with other macromolecules (372), allosteric effects by impacting distal residues from the phosphorylation site (373), and induction of conformational changes (374,375). Moreover, some conformational switches have been observed in numerous IDPs, in which threonines were implicated (365,376,377). One of these examples of IDPs with conformational changes resulting from phosphorylation, as well as its implication to function, is the 4E-BP2/eIF4E system. Briefly, 4EBP2 is an IDP that binds tightly to eIF4E to produce the inhibition of the cap-dependent translation initiation (367), which is one of the main mechanisms of ribosomal recruitment. Ribosomal recruitment of the mRNA is a key step in protein synthesis and also a target for the translational control in proteins (378). However, upon the phosphorylation of two threonine residues (Thr37 and Thr46), there is an important reduction in the affinity of 4EBP2 for eIF4E, causing eIF4G to compete with and displace 4E-BP2 from the eIF4E complex, therefore producing translation initiation (367).

The Thr68 residue, together with the polypeptide region around the Ala33, constitutes the recognized "hot spot" region within NUPR1. As previously mentioned, this region plays a crucial role in the interactions of NUPR1 with its natural biological partners (297,298,319,337), as well as with synthetic compounds and macromolecules (180,296,363). Moreover, given that the affinity of NUPR1 for its partners remains relatively similar in the assays carried out to date (296–298,319,337,363), it is possible that the phosphorylation at Thr68 partially influences the binding characteristics of NUPR1, in conjunction with the modulation produced by the other "hot spot" around Ala33. Notably, it is already known that Thr68 also participates in the binding of potent effective therapeutic agents targeting PC in mice (180,331). Therefore, the molecular effects induced by these drugs might be linked to conformational changes produced in NUPR1, thereby hampering its interactions with other natural partners and consequently disrupting the protein cascades in which NUPR1 plays a key role.

In order to summarize the first section of results comprising the functional characterization of NUPR1 in the binding to its partners, the following conclusions were obtained. Firstly, the interaction between the NLS region of NUPR1 and Impα3 was described by using a set of peptides comprising the specific mentioned polypeptide segment. This binding with Impα3 was influenced by the charges of K64 and K69 residues, but the most significant factor was the phosphorylation at Thr68. Importantly, Thr68 serves as a “hot spot” for many interactions of

NUPR1, making these findings a potential way for modulating the binding to its partners by targeting this residue. Additionally, from these results we can envisage a plausible mechanism for the action of drugs targeting NUPR1, as these drugs also bind via the Thr68. Therefore, understanding how NUPR1 is translocated to the nucleus, apart from providing more information on molecular mechanisms in the area of basic research, is of utmost utility in identifying new possible approaches to block the action of this protein, such as searching for active compounds that inhibit its binding to Impα, and thus preventing NUPR1 from exerting its functions. In conclusion of this part, and considering these results collectively, it was confirmed that NUPR1 requires the classical molecular mechanism of nuclear transport. Of note, it has not yet been clarified whether the NLS of a cargo first binds to Impα, thus releasing the autoinhibitory IBB domain for binding to Impβ; or whether upon interaction of Impα and Impβ, this binding displaces the IBB domain so that Impα can bind the NLS of the cargo (324). Thus, at this time there are two models of this mechanism of action, which need to be further studied. In this way, to enter NUPR1 into the nucleus: i) the NLS sequence of NUPR1 would bind to the ARM repeats of Impα, releasing the IBB domain, then binding the Impα/NUPR1 to Impβ, and this complex together would cross the NPC; or, alternatively, ii) Impα and Impβ would interact, displacing the IBB domain, and thus binding the NLS of the NUPR1 to Impα, to cross the NPC. Once inside the nucleus, this ternary complex must be dissociated by Ran-GTP, thus leaving NUPR1 in the nucleus, where it would carry out its multiple key cellular functions.

On the other hand, a functional characterization of the NUPR1L protein, a paralogue of NUPR1, was also carried out in this Doctoral Thesis, since NUPR1L is also an IDP that must be translocated to the cell nucleus to carry out its functions. Therefore, this characterization was performed in a very similar way to that performed in the case of NUPR1, in which the interaction of NUPR1L with Impα3, as the responsible protein for its nuclear translocation, was evaluated.

The first step was to confirm and validate the existence of the NLS region previously predicted for NUPR1L by docking simulation assays (data not shown, (336)). These results were obtained by our collaborators of this project, and they indicated that the NLS sequence of NUPR1L occupied the same region as its paralog NUPR1 (288,292). This NUPR1L region emerged as a key binding “hot spot”, establishing a connection with the major NLS-binding site on Impα3, situated within the ARM repeats 2-4. Our collaborators also demonstrated, by fluorescence, far-UV CD and NMR spectra, that the isolated NLS-NUPR1L lacked any propensity to adopt helix- or turn-like conformations (data not shown, (336)). Therefore, the NLS-NUPR1L behaved analogously to other NLS regions observed in well-folded proteins, meaning that this region remained disordered both in isolation and when participating in the formation of the complexes with importins (321,325,328,359).

Secondly, the inhibitory effect of the IBB domain in Impα3 was also evaluated in the case of the NUPR1L protein. The ITC results indicated that the removal of the IBB domain from Impα3, i.e., in the ΔImpα3 protein, enhanced the binding affinity of the NLS-NUPR1L peptide to the ARM 2-3 repeats of importins. Concretely, in the case of the NLS-NUPR1L peptide the K_d values were 12.0 μM for Impα3 while 5.5 μM for ΔImpα3. Therefore, the isolated region presented nearly equivalent binding affinities to both Impα3 and ΔImpα3 in comparison with the entire NUPR1L (K_d of 4.0 μM for Impα3, and K_d of 5.0 μM for ΔImpα3). Of note, these K_d values were obtained by fluorescence, not using ITC (data not shown, (336)), thus indicating that the essential amino acids, crucial for the binding, primarily resided within this specific polypeptide patch. Regrettably, a precise conclusion for the intact NUPR1L was not established due to the unavailability of binding parameter measurements using ITC. Nevertheless, these findings were in agreement with the previous results concerning other NLS regions of well-folded proteins (359,379,380), the intact NUPR1 protein (with a K_d value of 1.4 μM for Impα3 and 0.44 μM for ΔImpα3, as previously detailed), and in the case of peptides encompassing the NLS region of NUPR1 (previously detailed too). As in NUPR1, the presence of the IBB domain consistently imposed an autoinhibitory effect, hampering the effective anchoring of NLS-NUPR1L within the major NLS-binding region of Impα3. Moreover, these results indicated that importins had the capacity to interact with both paralogues, NUPR1L and NUPR1, with high affinities in the low micromolar range in both cases. Furthermore, as in the functional characterization of NUPR1, these results differed to those obtained recently in fluorescence-based techniques, where it was proposed that the nuclear translocation of other IDPs do not required the nucleus-cytoplasm transport mechanism (362), therefore reaffirming the fact that some IDPs need this classical mechanism of import.

On the other hand, since the isolated NLS regions of both NUPR1 and NUPR1L contained the essential residues required for interacting with Impα3, and considering that this binding using the peptides could be quantified through ITC, the binding affinities of these two paralogues were compared. By comparing the K_d values for NLS-NUPR1L (12.0 μM for Impα3 and 5.5 μM for ΔImpα3) with those of the NLS region of NUPR1 (1.7 μM for Impα3 and 0.95 μM for ΔImpα3), it was concluded that the binding affinity was stronger in the case of NUPR1. This discrepancy in the binding affinities for the same molecules between the two paralogues indicated that, despite they interacted with the same partners, their affinities for these partners differed. This difference could be potentially related to a mechanism of the regulation between these two proteins, being this regulation not only at the DNA level, but also at subsequent post-translational stages (319,320). Additionally, for the intact NUPR1 and NUPR1L proteins, it is worth remembering that a complete set of K_d values were not obtained using the same methodology, therefore being crucial to take into account the differences in the experimental techniques employed. Concretely, K_d values for NUPR1L were obtained by fluorescence, whereas for NUPR1, they were determined using ITC.

Therefore, as a conclusion of these last results, it was verified that NUPR1L is translocated to the cell nucleus through the classical mechanism of nuclear import, i.e., in the same way as NUPR1 does.

In parallel to the functional characterization of NUPR1 and NUPR1L, a small compound derived from TFP, known as ZZW-115, was developed in our group (331). ZZW-115 compound interacted with NUPR1 and acted as a potent NUPR1 inhibitor, exhibiting robust antitumor effects in PDAC as well as in other cancers (270,292,331). Due to the fact that ZZW-115 was a potent candidate for the treatment of PC, and due to the following aspects: i) the NLS sequence of NUPR1 played a crucial role in its interaction with Impα3 and subsequent nuclear translocation, ii) previous results of our collaborators established the binding of NUPR1 to the ZZW-115 compound via the residues Ala33 and Thr68 (331), and iii) Thr68 is located within the NLS of NUPR1 (331), it was assessed whether ZZW-115 had the capability to hamper the interaction between NUPR1 and Impα3. In this context, the mechanisms by which the ZZW-115 compound hampered the nuclear functions of NUPR1 were evaluated. Specifically, the finding obtained by out collaborators revealed that ZZW-115 interacted with NUPR1 through the residue Thr68 located within its predicted NLS sequence (292).

In the ITC assays, Impα3 exhibited a slightly higher affinity for NUPR1 (K_d value of 1.4 µM) compared to the affinity of NUPR1 for ZZW-115 (K_d value of 2.1 µM, (331)). However, the binding between Impα3 and NUPR1 showed a notably larger ΔH (-13.7 kcal/mol), compared to the ΔH of the binding between NUPR1 and ZZW-115 (-0.4 kcal/mol) (331). These differences in ΔH values probably indicated that a greater number of interactions between NUPR1 and Impα3 (60 kDa) occurred, compared to those with ZZW-115 (a smaller molecule). In the presence of ZZW-115 compound, the affinity of Impα3 for NUPR1 significantly decreased (K_d value of 35.0 µM at 100 µM ZZW-115), therefore producing a 25-fold reduction in the affinity. Despite the fact that Impα3 and ZZW-115 had a similar affinity for NUPR1, ZZW-115 changed the binding equilibrium between these two biomolecules likely through the following mechanisms: (i) the reduced size of ZZW-115, i.e., the binding of ZZW-115 to NUPR1 would be kinetically favored over the binding of NUPR1 to Impα3; (ii) ZZW-115 could be present at a similar or higher concentration than Impα3; and (iii) if the potential ternary complex NUPR1/ZZW-115/Impα3 was deficient in terms of nuclear translocation. These assays were carried out in conjunction with another series of techniques, performed by our collaborators, to verify in a more robust manner the inhibition that the compound ZZW-115 produced in the binding of NUPR1 to Impα3. For instance, in the absence of ZZW-115 it was observed that NUPR1 could be internalized into the nucleus, whereas in the presence of ZZW-115 it remained in the cytoplasm. Additionally, it was observed in MiaPaca-2 cells by PLA that the interaction of NUPR1 and Impα3 was considerably reduced in the presence of the compound ZZW-115 (data not shown, (292)). These findings provide further *in cellulo* evidence for the inhibitory role of ZZW-115, confirming that ZZW-115 interferes with NUPR1 internalization by interfering with the NUPR1-Impα3 interaction, and that NUPR1 requires Impα3 to be translocated to the nucleus. Collectively, these results demonstrated that ZZW-115 compound specifically bound to the NLS region of NUPR1 with an affinity comparable to that of importins for the protein, thereby partially modifying the binding equilibrium between NUPR1 and Impα3, i.e., it induced a cargo-specific displacement of importins. Consequently, ZZW-115 acted as an inhibitor of NUPR1 functions, hampering the nuclear translocation of NUPR1 and the protein-protein interaction of NUPR1 with other partners. Therefore, a screening approach to identify new high-affinity compounds utilizing the NLS-specific cargos as targets emerged as a promising strategy to use in future screening projects, which could effectively identify new targets of nuclear proteins, in agreement also with previous findings (381–383).

Unfortunately, during the initial pharmacological characterization of ZZW-115, it was discovered that this compound bound and blocked the hERG channel at low concentrations, and this interaction could disrupt cardiac repolarization and increase the risk of fatal arrhythmias, thus potentially causing cardiotoxicity in humans (334,335). Notably, while this side effect is observed in approximately 60% of drugs, it currently serves as an exclusion criterion for their clinical use in humans (332,333). Therefore, its development in further assays as a potent NUPR1 inhibitor was stopped. Consequently, a new screening of small compounds using another chemical library, the Maybridge library, was performed in this Doctoral Thesis, verifying the efficiency and the potential of the screening system of our research group.

As mentioned in the Introduction of Chapter II, several studies revealed that NUPR1 primarily exists in an unfolded state (298,337). Consequently, conventional methods for screening and identifying inhibitors cannot be directly employed for unfolded proteins such as NUPR1. Nevertheless, when a ligand interacts with a protein, even one in an unfolded state, this interaction could induce limited structural rearrangements, which may lead to different thermal denaturation patterns compared to the unliganded protein. Such alterations can be monitored using fluorescence and employing an extrinsic probe such as ANS. Thus, it is possible to use ligand-induced stabilization against thermal denaturation (TSA) methods to identify potential compounds that interact with NUPR1 (338). Concretely, the ANS extrinsic probe is able to bind to both hydrophobic patches or hydrophobic residues of proteins (214). Consequently, an increase in fluorescence intensity is generally observed during protein unfolding. However, this fluorescence intensity might decrease upon protein binding, depending on the changes in hydrophobicity in the surface area exposed to the solvent of a protein or protein complex. In this context, an exhaustive HTS of 10,000 compounds by a fluorescence thermal-denaturation method to discover compounds that interact with NUPR1 was performed, resulting in the identification of 45 potential candidates (AJO1-AJO45 compounds), which were considered as potential candidate compounds for further analysis. Subsequently, a set of techniques were carried out to evaluate the 45 compounds identified in the HTS.

Firstly, the interaction of the AJO compounds with NUPR1 was verified by ITC. Of these 45 compounds, AJO5, AJO17 and AJO22 showed solubility-related problems, so their binding affinities for NUPR1 were not evaluated by ITC. Of the remaining 42 AJO compounds, AJO14 showed a high affinity for the protein in the micromolar range (K_d value of 4.9 μM). Subsequently, the ability of these 42 compounds to have a suppressive growth effect was evaluated in MiaPaCa-2 cells. The compounds AJO14, AJO23 and AJO40 showed a high antiproliferative activity (IC_{50} values of 16.0 μM, 18.1 μM and 7.3 μM, respectively), therefore being AJO40 the best one. Of note, the compound ZZW-115 was used as a reference compound (IC_{50} value of 2.3 μM), to compare the IC_{50} values obtained with these compounds with those of a potent cell growth inhibitor. As can be seen, the IC_{50} values were higher than that of ZZW-115, so that in general, the potential inhibitors obtained in this HTS were in principle a bit less potent in exerting anticancer activity. Of all of them, the compound AJO40 initially generated more interest as it was the most effective compound, since, although its IC_{50} was higher than that of ZZW-115, it was not much higher, and it could have interesting important inhibitory properties.

It was then assessed if the anticancer effect observed with the compounds AJO14, AJO23 and AJO40, was indeed due to a NUPR1-related mechanism. To this end, Panc-1 cell line clones, concretely both NUPR1-WT and NUPR1-KO clones, were treated with these three compounds. The results revealed that the anticancer efficacy of AJO14 on NUPR1-WT cells was notably 2-3 times greater compared to NUPR1-KO cells (IC_{50} values of 8.2 μM in one WT clone, and 14.9 μM and 13 μM in two KO clones), indicating therefore that this compound exerted a direct effect on the NUPR1 protein. In contrast, AJO23 displayed comparable antiproliferative effects on both

NUPR1-WT and NUPR1-KO cells (IC_{50} values of 8.9 µM in one WT clone, and 8.4 µM and 7 µM in two KO clones), indicating that AJO23 lacked specificity for NUPR1 and must follow an alternative mechanism to reduce cell proliferation. Regrettably, despite its potent ability to decrease cell growth in MiaPaCa-2 cells (IC_{50} value of 7.3 µM), the results obtained in Panc-1 clones exhibited that KO clones showed a reduced viability compared to WT clones when exposed to AJO40 (IC_{50} values of 12.8 µM and 16 µM in two WT clones, and 10.9 µM and 8.8 µM in two KO clones). Consequently, it was evident that the efficacy of AJO40 was not specific to NUPR1. Importantly, at this point a search was performed for other possible results of assays carried out with the compound AJO40. The existence of a patent (US number 8,053,477 B2 (384)) was found, which described a series of compounds, including AJO40, used as inhibitors of the interaction between the proteins S100 and p53, and which activated the tumor suppressor activity of p53. Thus, the compounds described in this patent had an antineoplastic effect, and their use in the treatment of cancer was patented. Briefly, p53 is a transcription activator, triggering the upregulation of genes responsible for controlling cell cycle arrest and apoptosis, which plays a crucial role in the maintenance and regulation of the correct functioning of the cellular processes. When p53 presents mutations or is expressed at excessive levels, problems arise such as uncontrolled cell cycle unregulated and increased cell growth, as well as phenotypes associated with aging (385). Moreover, the regulation of p53 involves PTMs, including interactions with several proteins, such as some proteins of the S100 protein family, which significantly reduce its protein level and activity. In this context, S100 proteins might play a role in contributing to cell proliferation by binding to the C-terminus of p53, thereby hampering its function as a tumor suppressor (386,387). Therefore, and despite the substantial potency of AJO40 as inhibitor, its lack of specificity toward the target protein, in conjunction with its already known mechanism of action together with S100, produced its exclusion for further assays. Hence, AJO14 was the main candidate for exerting its anticancer effects via specifically targeting NUPR1. As a result, AJO14 was chosen as the most effective and potent compound with inhibitory properties for subsequent assays.

Subsequently, the antitumor activity of AJO14 was re-evaluated in primary cells derived from PDAC patients, which therefore offered a greater cellular and genetic variety, with different characteristics. In all of them, it was confirmed that this compound was able to reduce cell growth, and IC_{50} values in a range of 10.0 µM to 22.2 µM were obtained. The antitumor activity of AJO14 was also evaluated in different cancer cell lines, since, as previously mentioned, NUPR1 is overexpressed not only in PC, but also in a wide variety of cancers of different origin (304,346–354). The results obtained indicated that the compound AJO14 did not inhibit cell growth in only a particular cell type, since, in all the different cell lines treated with this compound, cell death was induced with IC_{50} values ranging from 10.3 µM to 22.0 µM. These findings confirmed again that NUPR1 is a good therapeutic target not only for treating PDAC but also for other type of cancers, being its targeting a promising strategy for cancer therapy.

An important aspect that has been commented on several occasions is the fact that, in order for a drug to continue its development towards the clinical phases in humans, it is necessary that, among other requirements, it does not bind and block the hERG channel, so as not to be a potential cause of cardiotoxicity in humans in the subsequent assays. Since the binding to this channel by the ZZW-115 compound halted its development towards the clinic, our group has established as a necessary and mandatory step, the measuring, with each compound identified as a potent drug in a project, whether it has a potential inhibitory effect on the hERG channel. In this way, the aim is to make an early and as correct as possible identification of potential compounds with anticancer activities, minimizing future side effects, as well as ensuring that subsequent assays, especially those involving animal models, are carried out in the most appropriate way possible without wasting unnecessary resources or animal suffering. In order to evaluate this possible binding between AJO14 and the hERG channel, a commercial kit called PredictorTM hERG fluorescence polarization assay was used, obtaining results that demonstrated that the IC_{50} values for E-4031 and AJO14 were 0.06 µM and higher than 80.0 µM, respectively. Therefore, revealing that the potency of AJO14 in suppressing the hERG channel *in vitro* was more than 1,000 times weaker when compared to E-4031. Thus, the probability for AJO14 causing cardiotoxicity associated with hERG binding was negligibly low, and could therefore be further advanced for subsequent preclinical studies.

The next step was to determine the molecular mechanisms by which AJO14 was able to carry out its antitumor activity. For this purpose, three commercial kits were used, by means of which specific parameters, which will be detailed below, could be evaluated *in cellulo* in MiaPaCa-2 cells. Firstly, the CytoTox-ONETM homogeneous membrane integrity assay kit was used to evaluate changes in LDH release from MiaPaCa-2 cells treated with AJO14. An increase in LDH release has been found to be associated with necrosis-mediated cell death (388). Necrosis is a type of cell death characterized by organelle swelling, membrane rupture, and subsequent plasma membrane breakdown, which results in the release of intracellular contents into the extracellular environment, triggering inflammation and, thus, inflammatory diseases (389,390). Initially, necrosis was considered as a passive and accidental form of cell death, in contrast to apoptosis, but later, necrosis has been proven to sometimes be a programmed process (391). The results showed that treatment with AJO14 induced a significant increase in this release from a concentration of 10 µM and above. At higher concentrations, around 60-80 µM, the LDH release produced increased considerably. These results were evident from 24 h of treatment with AJO14, with LDH release increasing over time (from 24 h to 72 h). Secondly, the Caspase-Glo® 3/7 Assay kit was used to evaluate changes in caspase-3/7 activity in MiaPaCa-2 cells treated with AJO14. An increase in caspase-3/7 activity is associated with apoptosis-mediated cell death (392). Apoptosis is known as a key form of programmed cell death, where genetically determined cell elimination takes place. It is a natural process during development and aging, serving as a homeostatic mechanism for regulating cell populations in tissues. Additionally, apoptosis acts as a defense mechanism in immune reactions or in response to cellular damage caused by diseases or harmful agents (393). The results indicated that treatment with AJO14 induced a significant increase in caspase-3/7 activity starting at a concentration of 5 µM. At higher concentrations, around 60-80 µM, the caspase-3/7 activity detected by this kit increased considerably, this result being visible as early as 24 h of treatment. These findings collectively indicated that the AJO14 compound promoted programmed cell death exerting pro-necrotic and apoptotic effects. Thirdly, the CellTiter-Glo® luminescent cell viability assay kit was used to evaluate changes at the level of ATP production and ATP release by MiaPaCa-2 cells treated with AJO14. Cellular ATP levels are vital for diverse biological processes and can be impaired in conditions like cancer and mitochondrial disorders. Mitochondria, known as cell powerhouses, produce ATP through oxidative phosphorylation. Besides energy, they play essential roles in stress responses, diseases, cell signaling, gene regulation, cell death, and many other functions (394,395). A decrease in ATP production and release is related to a functional failure of the mitochondria, being ATP necessary for the correct development of cellular functions (395). The results showed that treatment with AJO14 induced a significant decrease in cellular ATP content starting at 10 µM concentration at 24 h and 48 h, and 5 µM at 72 h. Therefore, these findings indicated that AJO14 treatment induced a cell metabolic failure in MiaPaCa-2 cells characterized by a decrease in ATP production.

Once elucidated the *in vitro* molecular mechanisms by which this antitumor activity was exerted by the AJO14 compound, its *in vivo* anticancer efficacy using a MiaPaCa-2 cell xenograft mouse model was evaluated. Of note, the primary *in vivo* systems employed in drug discovery involve xenograft models, where established immortalized cancer cell lines are injected into immunocompromised mice, with nude mice being a commonly used type because of their reduced T cell number due to a severely deteriorated or absent thymus. These models enable easy tracking of tumor growth along with initial studies on pharmacokinetics and pharmacodynamics of drug candidates. Xenografts offer reproducibility and efficiency as cultured cells are grown *in vitro*, collected and then uniformly injected into genetically similar animals. Due to its high reproducibility, low investment of time and resources, and ease of isolating tumor masses, xenografts are a widely integrated and frequently used mouse model in drug discovery (396,397). Concretely, in the assays performed in mice, during 28 days of treatment of the animals with AJO14, the tumor volumes and the weight of the mice were evaluated. In the case of mice from the control group, a significant increase in tumor size was observed (222.0 to 1,589.2 mm^3), compared to the tumor size of the mice in the groups treated with different concentrations of AJO14 (5 mg/kg: 269.7 to 1,130 mm^3; 10 mg/kg: 222.5 to 933.3 mm^3; 20 mg/kg: 230.8 to 858.2 mm^3; and 50 mg/kg: 246.4 to 225 mm^3). Moreover, regarding these data, the tumor growth arrest was dependent on the concentration of AJO14 used in each group of mice. Especially, in the case of mice treated with 50 mg/kg AJO14, almost no tumor growth was observed throughout the 28 days of the assay. On the other hand, the administration of AJO14 to mice, independently of the amount used, did not result in any decrease in body weight of the mice (control: 21.35 to 27.32 g; 5 mg/kg: 22.25 to 27.13 g; 10 mg/kg: 20.83 to 25.90; 20 mg/kg: 19.84 to 25.18 g; and 50 mg/kg: 20.42 g to 27.28 g). Therefore, the growth arrest on the tumors was dependent on AJO14 administration, and importantly, AJO14 could disrupt tumor growth without indications of systemic toxicity and any obvious neurological side effects in the mice.

In brief, all these results together indicated that the AJO14 compound was the most promising candidate, which exhibited a high specificity in targeting NUPR1, as demonstrated by the biophysical data obtained through ITC, as well as the notable insensitivity that NUPR1-KO cells showed, compared to NUPR1-WT cells, after the treatment with this compound. Furthermore, AJO14 displayed greater cytotoxic effects compared to the other AJO compounds, against PC cells and other cancer cell types *in vitro*. Moreover, AJO14 exhibited very low hERG affinity, thereby minimizing the risk of cardiotoxicity in future assays in clinical trials. It is also noteworthy that AJO14 promoted cell death through diverse mechanisms, involving necrosis, apoptosis, and an intense mitochondrial catastrophe, triggered by a reduction in ATP production with incompatible intracellular levels for cell survival. Surprisingly, all these functional features were similar than those observed with the ZZW-115, and through genetic inactivation of NUPR1 by specific NUPR1-targeting siRNA. Thereby indicating that AJO14 functioned by binding and inactivating the NUPR1 protein. The fact that the same compound induces cell death through different pathways and molecular mechanisms is of great interest, because tumor cells have to fight harder to surpass those effects produced by that compound. Therefore, the inhibitors that have been discovered and continue to be discovered in our research group are potential drugs and interesting compounds, as they do not exert their action through a single pathway. These results also indicate that NUPR1 is a very good target to combat PC. Furthermore, the results obtained *in vivo* in PDAC MiaPaCa-2 xenograft mice demonstrated the capacity of AJO14 to effectively regress tumor growth without inducing any negative side effect. Of note, while it was successfully demonstrated the specific NUPR1-targeting of AJO14, the potential existence of undetectable off-target effects associated with this compound must be evaluated in future assays. As a general conclusion, AJO14 represents a new potent NUPR1 inhibitor to treat PDAC without inducing cardiotoxicity.

Finally, another important aspect is that drugs found during the screening can be improved by introducing small modifications to obtain more effective and safer compounds. In this context, 51 derived-AJO14 compounds have been recently designed, synthesized and obtained by two methods to enhance both anticancer efficacy and physicochemical properties. One of these methods was based on conventional in-house laboratory design and the other one was a custom service performed by EDELRIS (Lyon, France), which included both a rational design of the drugs and an AI-based approach. Of note, using AI techniques for small-molecule drug discovery has attracted the attention of many drug discovery researchers in the past few years (398,399). AI-based tools can improve the predictability, precision, cost-effectiveness and the speed of drug discovery (400). Currently, all the methodology and experimental strategy employed with the AJO14 compound is being used with these new compounds in order to identity variants more efficient than their precursor AJO14 compound.

6. CONCLUSIONS

The following **conclusions** are derived from **Chapter II - Drug discovery for inhibiting NUPR1: from a chemical screening to an animal model**:

1. NUPR1 interacts with Impα3 to be translocated to the cell nucleus, thus requiring the classical molecular mechanism of protein nuclear transport.

2. The NLS region of NUPR1 is responsible for the interaction with Impα3.

3. The IBB domain of Impα3 exerts an autoinhibitory effect, hampering the effective anchoring of NLS-NUPR1 within the major NLS-binding site of Impα3.

4. The interaction of NUPR1 with Impα3 is influenced by the charges of K64 and K69 residues, but the most significant factor is the phosphorylation at Thr68.

5. The paralogue NUPR1L also interacts with Impα3 through the NLS to be translocated to the cell nucleus, as observed for NUPR1.

6. NLS-NUPR1 and NLS-NUPR1L behave analogously to other NLS regions observed in well-folded proteins, indicating that some IDPs need the classical molecular mechanism of protein nuclear transport.

7. Forty-five compounds (AJO compounds), identified by a high-throughput screening procedure, target NUPR1.

8. AJO14, the most promising compound, shows high affinity for NUPR1 with K_d values in the micromolar range, and displaying *in vitro* cytotoxic effects against PC cells and other cancer cell types.

9. AJO14 promotes cell death through necrosis, apoptosis, and an intense mitochondrial catastrophe, triggered by a reduction in ATP production.

10. AJO14 shows very low hERG affinity, minimizing the risk of cardiotoxicity in future clinical trials.

11. AJO14 inhibits tumor growth in PDAC MiaPaCa-2 xenograft mice, without inducing any *in vivo* negative side effect.

7. CONCLUSIONES

Las siguientes **conclusiones** se derivan del **Capítulo II - Descubrimiento de fármacos para inhibir NUPR1: de un cribado químico a un modelo animal**:

1. NUPR1 interactúa con Impα3 para ser translocada al núcleo celular, requiriendo así el mecanismo molecular clásico de transporte nuclear de proteínas.

2. La región NLS de NUPR1 es responsable de la interacción con Impα3.

3. El dominio IBB de Impα3 ejerce un efecto autoinhibitorio, dificultando el anclaje efectivo de NLS-NUPR1 dentro del sitio principal de unión a NLS de Impα3.

4. La interacción de NUPR1 con Impα3 está influenciada por las cargas de los residuos K64 y K69, pero el factor más significativo es la fosforilación en Thr68.

5. El parálogo NUPR1L también interactúa con Impα3 a través del NLS para ser translocado al núcleo celular, como se observó para NUPR1.

6. NLS-NUPR1 y NLS-NUPR1L se comportan de forma análoga a otras regiones NLS observadas en proteínas bien plegadas, lo que indica que algunas IDPs necesitan el mecanismo molecular clásico de transporte nuclear de proteínas.

7. Cuarenta y cinco compuestos (compuestos AJO), identificados mediante un procedimiento de cribado de alto rendimiento, se dirigen contra NUPR1.

8. AJO14, el compuesto más prometedor, muestra alta afinidad por NUPR1 con valores de K_d en el rango micromolar, y mostrando efectos citotóxicos *in vitro* contra células de CP y otros tipos de células cancerosas.

9. AJO14 promueve la muerte celular mediante necrosis, apoptosis y una intensa catástrofe mitocondrial, desencadenada por una reducción en la producción de ATP.

10. AJO14 muestra una afinidad hERG muy baja, lo que minimiza el riesgo de cardiotoxicidad en futuros ensayos clínicos.

11. AJO14 inhibe el crecimiento tumoral en ratones xenoinjertados de PDAC MiaPaCa-2, sin inducir ningún efecto secundario negativo *in vivo*.

REFERENCES

1. Pina AS, Hussain A, Roque ACA. An historical overview of drug discovery. Methods Mol Biol. 2009;572(1):3-12.

2. Dickson M, Gagnon JP. Key factors in the rising cost of new drug discovery and development. Nat Rev Drug Discov. 2004;3(5):417-29.

3. Van Norman GA. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. JACC Basic to Transl Sci. 2016;1(3):170-9.

4. Xu H, Xu HY, Lin MZ, Wang W, Li ZM, Huang JJ, et al. Learning the drug target-likeness of a protein. Proteomics. 2007;7(23):4255-63.

5. Ha J, Park H, Park J, Park SB. Recent advances in identifying protein targets in drug discovery. Cell Chem Biol. 2021;28(3):394-423.

6. Haberkorn U. What is cancer? Adv Nucl Oncol. 2007;62(4):1-16.

7. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359-86.

8. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Bannon F, et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population- based registries in 67 countries (CONCORD-2). Lancet. 2015;385(9972):977-1010.

9. Vargas-Román K, Tovar-Gálvez MI, Liñán-González A, Cañadas de la Fuente GA, de la Fuente-Solana EI, Díaz-Rodríguez L. Coping Strategies in Elderly Colorectal Cancer Patients. Cancers (Basel). 2022;14(3).

10. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66(4):683-91.

11. Ilic I, Ilic M. International patterns in incidence and mortality trends of pancreatic cancer in the last three decades: A joinpoint regression analysis. World J Gastroenterol. 2022;28(32):4698-715.

12. Black CJ, Drossman DA, Talley NJ, Ruddy J, Ford AC. Functional gastrointestinal disorders: advances in understanding and management. Lancet. 2020;396(10263):1664-74.

13. Singh R, Zogg H, Ghoshal UC, Ro S. Current Treatment Options and Therapeutic Insights

for Gastrointestinal Dysmotility and Functional Gastrointestinal Disorders. Front Pharmacol. 2022;13(January):1-20.

14. Holtmann G, Shah A, Morrison M. Pathophysiology of Functional Gastrointestinal Disorders: A Holistic Overview. Dig Dis. 2018;35(1):5-13.

15. Talley NJ. What Causes Functional Gastrointestinal Disorders? A Proposed Disease Model. Am J Gastroenterol. 2020;115(1):41-8.

16. Alfano CM, Molfino A, Muscaritoli M. Interventions to promote energy balance and cancer survivorship: Priorities for research and care. Cancer. 2013;119(SUPPL11):2143-50.

17. Gimeno-García AZ, Quintero E, Nicolás-Pérez D, Parra-Blanco A, Jiménez A. Colorectal cancer screening in a Spanish population. Med Clin (Barc). 2009;133(19):736-40.

18. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394(10207):1467-80.

19. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline From the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology. 2008;134(5):1570-95.

20. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi DJ, John A, et al. Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. Cancer. 2022;14(1732).

21. Mármol I, Sánchez-de-Diego C, Dieste AP, Cerrada E, Yoldi MJR. Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. Int J Mol Sci. 2017;18(1).

22. John SKP, George S, Primrose JN, Fozard JBJ. Symptoms and signs in patients with colorectal cancer. Color Dis. 2011;13(1):17-25.

23. Wolpin BM, Mayer RJ. Systemic Treatment of Colorectal Cancer. Gastroenterology. 2008;134(5):1296-310.

24. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: A review. JAMA - J Am Med Assoc. 2021;325(7):669-85.

25. Simon K. Colorectal cancer development and advances in screening. 2016;967-76.

26. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-Analyses of colorectal cancer risk factors. Cancer Causes Control. 2013;24(6):1207-22.

27. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer. 2006;118(12):3030-44.

28. Burnett-Hartman AN, Newcomb PA, Potter JD. Infectious agents and colorectal cancer: A review of Helicobacter pylori, Streptococcus bovis, JC virus, and human papillomavirus. Cancer Epidemiol Biomarkers Prev. 2008;17(11):2970-9.

29. Guarner F. Enteric flora in health and disease. Digestion. 2006;73(SUPPL. 1):5-12.

30. Falk PG, Hooper L V., Midtvedt T, Gordon JI. Creating and Maintaining the Gastrointestinal Ecosystem: What We Know and Need To Know from Gnotobiology. Microbiol Mol Biol Rev. 1998;62(4):1157-70.

31. Wexler HM. Bacteroides: The good, the bad, and the nitty-gritty. Clin Microbiol Rev. 2007;20(4):593-621.

32. Reid G. When microbe meets human. Clin Infect Dis. 2004;39(6):827-30.

33. Xu J, Gordon JI. Honor thy symbionts. Proc Natl Acad Sci U S A. 2003;100(18):10452-9.

34. Young VB. The role of the microbiome in human health and disease: An introduction for clinicians. BMJ. 2017;356.

35. Kling JJ, Wright RL, Moncrief JS, Wilkins TD. Cloning and characterization of the gene for the metalloprotease enterotoxin. FEMS Microbiol Lett. 1997;146:279-84.

36. Yekani M, Baghi HB, Naghili B, Vahed SZ, Sóki J, Memar MY. To resist and persist: Important factors in the pathogenesis of Bacteroides fragilis. Microb Pathog. 2020;149(April).

37. Goldstein EJC. Anaerobic bacteremia. Clin Infect Dis. 1996;23:97-101.

38. Jasemi S, Emaneini M, Ahmadinejad Z, Fazeli MS, Sechi LA, Sadeghpour Heravi F, et al. Antibiotic resistance pattern of Bacteroides fragilis isolated from clinical and colorectal specimens. Ann Clin Microbiol Antimicrob. 2021;20(1):1-8.

39. Prindiville TP, Sheikh RA, Cohen SH, Tang YJ, Cantrell MC, Silva J. Bacteroides fragilis enterotoxin gene sequences in patients with inflammatory bowel disease. Emerg Infect Dis. 2000;6(2):171-4.

40. Deng H, Li Z, Tan Y, Guo Z, Liu Y, Wang Y, et al. A novel strain of Bacteroides fragilis enhances phagocytosis and polarises M1 macrophages. Sci Rep. 2016;6(20):1-11.

41. Baughn AD, Malamy MH. The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen. Nature. 2004;427(6973):441-4.

42. Sears CL. Enterotoxigenic Bacteroides fragilis: A rogue among symbiotes. Clin Microbiol Rev. 2009;22(2):349-69.

43. Cheng WT, Kantilal HK, Davamani F. The mechanism of bacteroides fragilis toxin contributes to colon cancer formation. Malaysian J Med Sci. 2020;27(4):9-21.

44. Sears CL, Geis AL, Housseau F. Bacteroides fragilis subverts mucosal biology: From symbiont to colon carcinogenesis. J Clin Invest. 2014;124(10):4166-72.

45. Myers LL, Firehammer BD, Shoop DS, Border MM. Bacteroides fragilis: A possible cause of acute diarrheal disease in newborn lambs. Infect Immun. 1984;44(2):241-4.

46. Myers LL, Shoop DS, Stackhouse LL, Newman FS, Flaherty RJ, Letson GW, et al. Isolation of enterotoxigenic Bacteroides fragilis from humans with diarrhea. J Clin Microbiol. 1987;25(12):2330-3.

47. Valguarnera E, Wardenburg JB. Good Gone Bad: One Toxin Away From Disease for Bacteroides fragilis. J Mol Biol. 2020;432(4):765-85.

48. Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. Cancer J (United States). 2014;20(3):181-9.

49. Lin FR, Niparko JK, Ferrucci and L. Stat3 activation in murine colitis induced by enterotoxigenic Bacteroides fragilis. Bone. 2014;23(1):1-7.

50. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu XQ, Murray-Stewart TR, et al. Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. Proc Natl Acad Sci U S A. 2011;108(37):15354-9.

51. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med. 2009;15(9):1016-22.

52. Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, Lazarev MG, et al. The bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. Clin Infect Dis. 2015;60(2):208-15.

53. Rabizadeh S, Rhee KJ, Wu S, Huso D, Gan CM, Golub JE, et al. Enterotoxigenic Bacteroides fragilis: A potential instigator of colitis. Inflamm Bowel Dis. 2007;13(12):1475-83.

54. Rhee KJ, Wu S, Wu X, Huso DL, Karim B, Franco AA, et al. Induction of persistent colitis by a human commensal, enterotoxigenic Bacteroides fragilis, in wild-type C57BL/6 mice. Infect Immun. 2009;77(4):1708-18.

55. Palit P, Das R, Haque MA, Nuzhat S, Khan SS, Siddiqua TJ, et al. Risk Factors for Enterotoxigenic Bacteroides fragilis Infection and Association with Environmental Enteric Dysfunction and Linear Growth in Children: Results from the MAL-ED Study. Am J Trop Med Hyg. 2022;106(3):915-22.

56. Fathi P, Wu S. Isolation, Detection, and Characterization of Enterotoxigenic Bacteroides fragilis in Clinical Samples. Open Microbiol J. 2016;10(1):57-63.

57. Cohen SH, Shetab R, Tang-Feldman YJ, Sarma P, Silva J, Prindiville TP. Prevalence of enterotoxigenic Bacteroides fragilis in hospital-acquired diarrhea. Diagn Microbiol Infect Dis. 2006;55(4):251-4.

58. Miki T, Kuwahara T, Nakayama H, Okada N, Kataoka K, Arimochi H, et al. Simultaneous detection of Bacteroides fragilis group species by leuB-directed PCR. J Med Investig. 2005;52:93-100.

59. Lee CS, Lee J. Evaluation of new gyrB-based real-time PCR system for the detection of B. fragilis as an indicator of human-specific fecal contamination. J Microbiol Methods. 2010;82(3):311-8.

60. Tong J, Liu C, Summanen P, Xu H, Finegold SM. Application of quantitative real-time PCR for rapid identification of Bacteroides fragilis group and related organisms in human wound samples. Anaerobe. 2011;17(2):64-8.

61. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: Paradigm shift in the human gut microbiome study. Clin Microbiol Infect. 2012;18(12):1185-93.

62. Sears CL, Islam S, Saha A, Arjumand M, Alam NH, Faruque ASG, et al. Association of enterotoxigenic Bacteroides fragilis infection with inflammatory diarrhea. Clin Infect Dis. 2008;47(6):797-803.

63. Franco AA, Mundy LM, Trucksis M, Wu S, Kaper JB, Sears CL. Cloning and

characterization of the Bacteroides fragilis metalloprotease toxin gene. Infect Immun. 1997;65(3):1007-13.

64. Wu S, Dreyfus LA, Tzianabos AO, Hayashi C, Sears CL. Diversity of the metalloprotease toxin produced by enterotoxigenic Bacteroides fragilis. Infect Immun. 2002;70(5):2463-71.

65. Chung GT, Franco AA, Wu S, Rhie GE, Cheng R, Oh HB, et al. Identification of a third metalloprotease toxin gene in extraintestinal isolates of Bacteroides fragilis. Infect Immun. 1999;67(9):4945-9.

66. Avila-Campos MJ, Liu C, Song Y, Rowlinson MC, Finegold SM. Determination of bft gene subtypes in Bacteroides fragilis clinical isolates. J Clin Microbiol. 2007;45(4):1336-8.

67. Goulas T, Arolas JL, Gomis-Rüth FX. Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. Proc Natl Acad Sci U S A. 2011;108(5):1856-61.

68. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. Pathogenicity islands of virulent bacteria: Structure, function and impact on microbial evolution. Mol Microbiol. 1997;23(6):1089-97.

69. Franco AA, Cheng RK, Chung GT, Wu S, Oh HB, Sears CL. Molecular evolution of the pathogenicity island of enterotoxigenic Bacteroides fragilis strains. J Bacteriol. 1999;181(21):6623-33.

70. Moncrief JS, Duncan AJ, Wright RL, Barroso LA, Wilkins TD. Molecular characterization of the fragilysin pathogenicity islet of enterotoxigenic Bacteroides fragilis. Infect Immun. 1998;66(4):1735-9.

71. Franco AA. The Bacteroides fragilis pathogenicity island is contained in a putative novel conjugative transposon. J Bacteriol. 2004;186(18):6077-92.

72. Buckwold SL, Shoemaker NB, Sears CL, Franco AA. Identification and characterization of conjugative transposons CTn86 and CTn9343 in Bacteroides fragilis strains. Appl Environ Microbiol. 2007;73(1):53-63.

73. Claros MC, Claros ZC, Hecht DW, Citron DM, Goldstein EJC, Silva J, et al. Characterization of the Bacteroides fragilis pathogenicity island in human blood culture isolates. Anaerobe. 2006;12(1):17-22.

74. Moncrief JS, Obiso R, Barroso LA, Kling JJ, Wright RL, Van Tassell RL, et al. The enterotoxin of Bacteroides fragilis is a metalloprotease. Infect Immun. 1995;63(1):175-81.

75. Wu S, Shin J, Zhang G, Cohen M, Franco A, Sears CL. The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor. Infect Immun. 2006;74(9):5382-90.

76. Wu S, Lim KC, Huang J, Saidi RF, Sears CL. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. Proc Natl Acad Sci U S A. 1998;95(25):14979-84.

77. Massova I, Kotra LP, Fridman R, Mobashery S. Matrix metalloproteinases : structures, evolution, and diversification. FASEB J. 2017;12(12):1075-95.

78. Chambers FG, Koshy SS, Saidi RF, Clark DP, Moore RD, Sears CL. Bacteroides fragilis toxin exhibits polar activity on monolayers of human intestinal epithelial cells (T84 cells) in vitro. Infect Immun. 1997;65(9):3561-70.

79. Obiso RJ, Lyerly DM, Van Tassell RL, Wilkins TD. Proteolytic activity of the Bacteroides fragilis enterotoxin causes fluid secretion and intestinal damage in vivo. Infect Immun. 1995;63(10):3820-6.

80. Wu S, Morin PJ, Maouyo D, Sears CL. Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. Gastroenterology. 2003;124(2):392-400.

81. Wu S, Rhee KJ, Zhang M, Franco A, Sears CL. Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and γ-secretase-dependent E-cadherin cleavage. J Cell Sci. 2007;120(11):1944-52.

82. Nelson WJ, Nusse R. Convergence of Wnt, β-Catenin, and Cadherin pathways. Science (80-). 2004;303(5663):1483-7.

83. Tian X, Liu Z, Niu B, Zhang J, Tan TK, Lee SR, et al. E-Cadherin/β-catenin complex and the epithelial barrier. J Biomed Biotechnol. 2011;2011.

84. Devaux CA, Mezouar S, Mege JL. The E-Cadherin Cleavage Associated to Pathogenic Bacteria Infections Can Favor Bacterial Invasion and Transmigration, Dysregulation of the Immune Response and Cancer Induction in Humans. Front Microbiol. 2019;10(November):1-19.

85. Yu H, Jove R. The stats of cancer - New molecular targets come of age. Nat Rev Cancer. 2004;4(2):97-105.

86. Yu H, Kortylewski M, Pardoll D. Crosstalk between cancer and immune cells: Role of STAT3 in the tumour microenvironment. Nat Rev Immunol. 2007;7(1):41-51.

87. Sanfilippo L, Li CKF, Seth R, Balwin TJ, Menozzi MG, Mahida YR. Bacteroides fragilis enterotoxin induces the expression of IL-8 and transforming growth factor-beta (TGF-β) by human colonic epithelial cells. Clin Exp Immunol. 2000;119(3):456-63.

88. Zinner SH. Antibiotic use: Present and future. New Microbiol. 2007;30(3):321-5.

89. Singh SB, Barrett JF. Empirical antibacterial drug discovery - Foundation in natural products. Biochem Pharmacol. 2006;71(7):1006-15.

90. Laxminarayan R, Matsoso P, Pant S, Brower C, Røttingen JA, Klugman K, et al. Access to effective antimicrobials: A worldwide challenge. Lancet. 2016;387(10014):168-75.

91. Singh SB, Young K, Silver LL. What is an " Ideal " Antibiotic ? Discovery Challenges and Path forward. Biochem Pharmacol. 2017;(January).

92. Zaman S Bin, Hussain MA, Nye R, Mehta V, Mamun KT, Hossain N. A Review on Antibiotic Resistance: Alarm Bells are Ringing. Cureus. 2017;9(6).

93. Davies J. Origins and evolution of antibiotic resistance. Microbiologia. 1996;12(1):9-16.

94. Moussa SH, Tayel AA, Al-Hassan AA, Farouk A. Tetrazolium/Formazan Test as an Efficient Method to Determine Fungal Chitosan Antimicrobial Activity. J Mycol. 2013;2013:1-7.

95. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629-55.

96. Shankar Pr. Tackling drug-resistant infections globally: final report and recommendations. Arch Pharm Pract. 2016;7(3):110.

97. Neill JO'. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations The Review on Antimicrobial Resistance Chaired. 2014;(December).

98. Sóki J, Wybo I, Hajdú E, Toprak NU, Jeverica S, Stingu CS, et al. A Europe-wide assessment of antibiotic resistance rates in Bacteroides and Parabacteroides isolates from intestinal microbiota of healthy subjects. Anaerobe. 2020;62:102182.

99. Pumbwe L, Wareham DW, Aduse-Opoku J, Brazier JS, Wexler HM. Genetic analysis of

mechanisms of multidrug resistance in a clinical isolate of Bacteroides fragilis. Clin Microbiol Infect. 2007;13(2):183-9.

100. Kangaba AA, Saglam FY, Tokman HB, Torun M, Torun MM. The prevalence of enterotoxin and antibiotic resistance genes in clinical and intestinal Bacteroides fragilis group isolates in Turkey. Anaerobe. 2015;35:72-6.

101. Nagy E, Urbán E. Antimicrobial susceptibility of Bacteroides fragilis group isolates in Europe: 20 years of experience. Clin Microbiol Infect. 2011;17(3):371-9.

102. Fille M, Mango M, Lechner M, Schaumann R. Bacteroides fragilis group: Trends in resistance. Curr Microbiol. 2006;52(2):153-7.

103. Liu CY, Huang YT, Liao CH, Yen LC, Lin HY, Hsueh PR. Increasing trends in antimicrobial resistance among clinically important anaerobes and bacteroides fragilis isolates causing nosocomial infections: Emerging resistance to carbapenems. Antimicrob Agents Chemother. 2008;52(9):3161-8.

104. Snydman DR, Jacobus N V., McDermott LA, Ruthazer R, Golan Y, Goldstein EJC, et al. National survey on the susceptibility of Bacteroides fragilis group: Report and analysis of trends in the United States from 1997 to 2004. Antimicrob Agents Chemother. 2007;51(5):1649-55.

105. Goldstein EJC, Citron DM, Tyrrell KL. In vitro activity of eravacycline and comparator antimicrobials against 143 recent strains of Bacteroides and Parabacteroides species. Anaerobe. 2018;52(2018):122-4.

106. Treviño M, Areses P, Dolores Peñalver M, Cortizo S, Pardo F, Luisa Pérez del Molino M, et al. Susceptibility trends of Bacteroides fragilis group and characterisation of carbapenemase-producing strains by automated REP-PCR and MALDI TOF. Anaerobe. 2012;18(1):37-43.

107. Maraki S, Mavromanolaki VE, Stafylaki D, Kasimati A. Surveillance of antimicrobial resistance in recent clinical isolates of Gram-negative anaerobic bacteria in a Greek University Hospital. Anaerobe. 2020;62.

108. Wang Y, Han Y, Shen H, Lv Y, Zheng W, Wang J. Higher prevalence of multi-antimicrobial resistant bacteroides spp. Strains isolated at a tertiary teaching hospital in China. Infect Drug Resist. 2020;13:1537-46.

109. Sárvári KP, Sóki J, Kristóf K, Juhász E, Miszti C, Melegh SZ, et al. Molecular

characterisation of multidrug-resistant Bacteroides isolates from Hungarian clinical samples. J Glob Antimicrob Resist. 2018;13(2010):65-9.

110. Cordovana M, Kostrzewa M, Sóki J, Witt E, Ambretti S, Pranada AB. Bacteroides fragilis: A whole MALDI-based workflow from identification to confirmation of carbapenemase production for routine laboratories. Anaerobe. 2018;54:246-53.

111. Lage OM, Ramos MC, Calisto R, Almeida E, Vasconcelos V, Vicente F. Current screening methodologies in drug discovery for selected human diseases. Mar Drugs. 2018;16(8):1-31.

112. Ericsson AC, Crim MJ, Franklin CL. A brief history of animal modeling. Mo Med. 2013;110(3):201-5.

113. Robinson NB, Krieger K, Khan F, Huffman W, Chang M, Naik A, et al. The current state of animal models in research: A review. Int J Surg. 2019;72(October):9-13.

114. De Boo J, Hendriksen C. Reduction strategies in animal research: A review of scientific approaches at the intra-experimental, supra-experimental and extra-experimental levels. ATLA Altern to Lab Anim. 2005;33(4):369-77.

115. Berman JN, Chiu PPL, Dellaire G. Preclinical Animal Models for Cancer Genomics. Cancer Genomics: From Bench to Personalized Medicine. Elsevier Inc.; 2013. 109-131 p.

116. Onderdonk AB. Animal models simulating anaerobic infections. Anaerobe. 2005;11(4):189-95.

117. Ségalat L. Invertebrate Animal Models of Diseases as Screening Tools in Drug Discovery. ACS Chem Biol. 2007;2(4):231-6.

118. Kelly MJ. Wound infection: A controlled clinical and experimental demonstration of synergy between aerobic (Escherichia coli) and anaerobic (Bacteroides fragilis) bacteria. Ann R Coll Surg Engl. 1980;62(1):52-9.

119. Kelly MJ. The quantitative and histological demonstration of pathogenic synergy between Escherichia coli and Bacteroides fragilis in guinea-pig wounds. 1978;11.

120. Ochoa-Repáraz J, Mielcarz DW, Wang Y, Begum-Haque S, Dasgupta S, Kasper DL, et al. A polysaccharide from the human commensal Bacteroides fragilis protects against CNS demyelinating disease. Mucosal Immunol. 2010;3(5):487-95.

121. Levy N. The use of animal as models: Ethical considerations. Int J Stroke. 2012;7(5):440-

2.

122. Tsai CJY, Loh JMS, Proft T. Galleria mellonella infection models for the study of bacterial diseases and for antimicrobial drug testing. Virulence. 2016;7(3):214-29.

123. Wilson-Sanders SE. Invertebrate models for biomedical research, testing, and education. ILAR J. 2011;52(2):126-52.

124. Pereira TC, Barros PP de, de Oliveira Fugisaki LR, Rossoni RD, Ribeiro F de C, Menezes RT de, et al. Recent advances in the use of Galleria mellonella model to study immune responses against human pathogens. J Fungi. 2018;4(4).

125. Cutuli MA, Petronio Petronio G, Vergalito F, Magnifico I, Pietrangelo L, Venditti N, et al. Galleria mellonella as a consolidated in vivo model hosts: New developments in antibacterial strategies and novel drug testing. Virulence. 2019;10(1):527-41.

126. Ménard G, Rouillon A, Cattoir V, Donnio PY. Galleria mellonella as a Suitable Model of Bacterial Infection: Past, Present and Future. Front Cell Infect Microbiol. 2021;11(December):1-18.

127. Brennan M, Thomas DY, Whiteway M, Kavanagh K. Correlation between virulence of Candida albicans mutants in mice and Galleria mellonella larvae. FEMS Immunol Med Microbiol. 2002;34(2):153-7.

128. Desalermos A, Tan X, Rajamuthiah R, Arvanitis M, Wang Y, Li D, et al. A multi-host approach for the systematic analysis of virulence factors in cryptococcus neoformans. J Infect Dis. 2015;211(2):298-305.

129. Brunke S, Quintin J, Kasper L, Jacobsen ID, Richter ME, Hiller E, et al. Of mice, flies - and men? Comparing fungal infection models for large-scale screening efforts. DMM Dis Model Mech. 2015;8(5):473-86.

130. Kwadha CA, Ong'Amo GO, Ndegwa PN, Raina SK, Fombong AT. The biology and control of the greater wax moth, Galleria mellonella. Insects. 2017;8(2):1-17.

131. Fuchs BB, O'Brien E, Khoury JBE, Mylonakis E. Methods for using Galleria mellonella as a model host to study fungal pathogenesis. Virulence. 2010;1(6):475-82.

132. Jorjão AL, Oliveira LD, Scorzoni L, Figueiredo-Godoi LMA, Prata MCA, Jorge AOC, et al. From moths to caterpillars: Ideal conditions for galleria mellonella rearing for in vivo microbiological studies. Virulence. 2018;9(1):383-9.

133. Ellis JD, Graham JR, Mortensen A. Standard methods for wax moth research. J Apic Res. 2013;52(1).

134. Charrière JD, Imdorf A. Protection of honey combs from wax moth damage. Am Bee J. 1999;139(8):627-30.

135. Stephen WP. Honey Bee Pests, Predators, and Diseases. Roger A. Morse . Vol. 55, The Quarterly Review of Biology. 1980. p. 193-193.

136. Powning RF, Davidson WJ. Studies on insect bacteriolytic enzymes-I. Lysozyme in haemolymph of Galleria mellonella and Bombyx mori. Comp Biochem Physiol -- Part B Biochem. 1973;45(3).

137. Firacative C, Khan A, Duan S, Ferreira-Paim K, Leemon D, Meyer W. Rearing and maintenance of galleria mellonella and its application to study fungal virulence. J Fungi. 2020;6(3):1-13.

138. Nathan S. New to galleria mellonella: Modeling an ExPEC infection. Virulence. 2014;5(3):371-4.

139. Wojda I. Immunity of the greater wax moth Galleria mellonella. Insect Sci. 2017;24(3):342-57.

140. Browne N, Heelan M, Kavanagh K. An analysis of the structural and functional similarities of insect hemocytes and mammalian phagocytes. Virulence. 2013;4(7):597-603.

141. Ramarao N, Nielsen-Leroux C, Lereclus D. The insect Galleria mellonella as a powerful infection model to investigate bacterial pathogenesis. J Vis Exp. 2012;(70):1-7.

142. Mukherjee K, Hain T, Fischer R, Chakraborty T, Vilcinskas A. Brain infection and activation of neuronal repair mechanisms by the human pathogen Listeria monocytogenes in the lepidopteran model host Galleria mellonella. Virulence. 2013;4(4):324-32.

143. Arteaga Blanco LA, Crispim JS, Fernandes KM, de Oliveira LL, Pereira MF, Bazzolli DMS, et al. Differential cellular immune response of Galleria mellonella to Actinobacillus pleuropneumoniae. Cell Tissue Res. 2017;370(1):153-68.

144. Boman HG, Hultmark D. Cell-free immunity in insects. Annu Rev Microbiol. 1987;41:103-26.

145. Pereira MF, Rossi CC, Da Silva GC, Rosa JN, Bazzolli DMS. Galleria mellonella as an infection model: An in-depth look at why it works and practical considerations for successful application. Pathog Dis. 2020;78(8):1-15.

146. Trevijano-Contador N, Zaragoza O. Immune response of Galleria mellonella against human fungal pathogens. J Fungi. 2019;5(1):1-13.

147. Brown SE, Howard A, Kasprzak AB, Gordon KH, East PD. A peptidomics study reveals the impressive antimicrobial peptide arsenal of the wax moth Galleria mellonella. Insect Biochem Mol Biol. 2009;39(11):792-800.

148. Vogel H, Altincicek B, Glöckner G, Vilcinskas A. A comprehensive transcriptome and immune-gene repertoire of the lepidopteran model host Galleria mellonella. BMC Genomics. 2011;12.

149. Vilcinskas A, Wedde M. Insect inhibitors of metalloproteinases. IUBMB Life. 2002;54(6):339-43.

150. Hoffmann JA, Reichhart JM, Hetru C. Innate immunity in higher insects. Curr Opin Immunol. 1996;8(1):8-13.

151. Cerenius L, Lee BL, Söderhäll K. The proPO-system: pros and cons for its role in invertebrate immunity. Trends Immunol. 2008;29(6):263-71.

152. Kopácek P, Weise C, Götz P. The prophenoloxidase from the wax moth Galleria mellonella: purification and characterization of the proenzyme. Insect Biochem Mol Biol. 1995;25(10):1081-91.

153. Melillo D, Marino R, Italiani P, Boraschi D. Innate Immune Memory in Invertebrate Metazoans: A Critical Appraisal. Front Immunol. 2018;9(August).

154. Pham LN, Dionne MS, Shirasu-Hiza M, Schneider DS. A specific primed immune response in Drosophila is dependent on phagocytes. PLoS Pathog. 2007;3(3).

155. Chen RY, Andrew Keddie B. The Galleria mellonella-Enteropathogenic Escherichia coli Model System: Characterization of Pathogen Virulence and Insect Immune Responses. J Insect Sci. 2021;21(4).

156. Binder U, Maurer E, Lass-Flörl C. Galleria mellonella: An invertebrate model to study pathogenicity in correctly defined fungal species. Fungal Biol. 2016;120(2):288-95.

157. Arvanitis M, Glavis-Bloom J, Mylonakis E. Invertebrate models of fungal infection.

Biochim Biophys Acta - Mol Basis Dis. 2013;1832(9):1378-83.

158. Panayidou S, Ioannidou E, Apidianakis Y. Human pathogenic bacteria, fungi, and viruses in Drosophila. Virulence. 2014;5(2):253-69.

159. López Hernández Y, Yero D, Pinos-Rodríguez JM, Gibert I. Animals devoid of pulmonary system as infection models in the study of lung bacterial pathogens. Front Microbiol. 2015;6(FEB):1-19.

160. Graham ML, Prescott MJ. The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease. Eur J Pharmacol. 2015;759:19-29.

161. Hubrecht RC, Carter E. The 3Rs and humane experimental technique: Implementing change. Animals. 2019;9(10):1-10.

162. Champion OL, Titball RW, Bates S. Standardization of G. mellonella larvae to provide reliable and reproducible results in the study of fungal pathogens. J Fungi. 2018;4(3).

163. Banville N, Browne N, Kavanagh K. Effect of nutrient deprivation on the susceptibility of Galleria mellonella larvae to infection. Virulence. 2012;3(6):497-503.

164. Junqueira JC. Galleria mellonella as a model host for human pathogens: Recent studies and new perspectives. Virulence. 2012;3(6):474-6.

165. Lange A, Schäfer A, Frick JS. A Galleria mellonella Oral Administration Model to Study Commensal-Induced Innate Immune Responses. J Vis Exp. 2019;(145).

166. Özkan S, Coutts RHA. Aspergillus fumigatus mycovirus causes mild hypervirulent effect on pathogenicity when tested on Galleria mellonella. Fungal Genet Biol. 2015;76:20-6.

167. Kay S, Edwards J, Brown J, Dixon R. Galleria mellonella infection model identifies both high and low lethality of clostridium perfringens toxigenic strains and their response to antimicrobials. Front Microbiol. 2019;10(JUL):1-11.

168. Pereira MF, Rossi CC, de Queiroz MV, Martins GF, Isaac C, Bossé JT, et al. Galleria mellonella is an effective model to study Actinobacillus pleuropneumoniae infection. Microbiol (United Kingdom). 2015;161(2):387-400.

169. Sheehan G, Dixon A, Kavanagh K. Utilization of galleria mellonella larvae to characterize the development of staphylococcus aureus infection. Microbiol (United Kingdom). 2019;165(8):863-75.

170. Bolouri Moghaddam MR, Tonk M, Schreiber C, Salzig D, Czermak P, Vilcinskas A, et al. The potential of the Galleria mellonella innate immune system is maximized by the co-presentation of diverse antimicrobial peptides. Biol Chem. 2016;397(9):939-45.

171. Organism M, Study T. Establishment and Validation of Galleria mellonella as a Novel Model Organism To Study Mycobacterium abscessus Infection, Pathogenesis, and Treatment. 2018;1-9.

172. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta - Proteins Proteomics. 2005;1751(2):119-39.

173. Hellmann N, Schneider D. Hands On: Using Tryptophan Fluorescence Spectroscopy to Study Protein Structure. Methods Mol Biol. 2019;1958:379-401.

174. Gao K, Oerlemans R, Groves MR. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. Biophys Rev. 2020;12(1):85-104.

175. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen. 2001;6:429-40.

176. Matulis D, Kranz JK, Salemme FR, Todd MJ. Thermodynamic stability of carbonic anhydrase: Measurements of binding affinity and stoichiometry using thermofluor. Biochemistry. 2005;44(13):5258-66.

177. Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem. 2004;332(1):153-9.

178. Johnson CM. Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys. 2013;531(1-2):100-9.

179. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;3(25).

180. Neira JL, Bintz J, Arruebo M, Rizzuti B, Bonacci T, Vega S, et al. Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. Sci Rep. 2017;7(June 2016):39732.

181. Abian O, Vega S, Sancho J, Velazquez-Campoy A. Allosteric Inhibitors of the NS3

Protease from the Hepatitis C Virus. PLoS One. 2013;8(7):1-10.

182. Cremades N, Hermoso J, Jime P, Lanas A, Œ PSH, Sancho J. Discovery of Specific Flavodoxin Inhibitors as Potential Therapeutic Agents against Helicobacter pylori Infection. ACS Chem Biol. 2009;4(11):928-38.

183. Pey AL, Ying M, Cremades N, Velazquez-Campoy A, Scherer T, Thöny B, et al. Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. J Clin Invest. 2008;118(8):2858-67.

184. Huynh K, Partch CL. Current Protocols in Protein Science: Analysis of protein stability and ligand interactions by thermal shift assay. Curr Protoc Protein Sci. 2016;79:28.9.1-28.9.14.

185. Velazquez-Campoy A, Sancho J, Abian O, Vega S. Biophysical Screening for Identifying Pharmacological Chaperones and Inhibitors against Conformational and Infectious Diseases. Curr Drug Targets. 2016;17:1492-505.

186. Bastos M, Abian O, Johnson CM, Ferreira-da-silva F, Vega S, Jiménez-Alesanco A, et al. Isothermal titration calorimetry. Nat Rev methods Prim. 2023;3(17).

187. Lewis EA, Murphy KP. Isothermal titration calorimetry. Methods Mol Biol. 2005;305(7):1-16.

188. Velzquez-Campoy A, Ohtaka H, Nezami A, Muzammil S, Freire E. Isothermal Titration Calorimetry. Curr Protoc Cell Biol. 2004;17(8):1-24.

189. Velazquez-Campoy A, Freire E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protoc. 2006;1(1):186-91.

190. Bastos M, Velazquez-Campoy A. Isothermal titration calorimetry (ITC): a standard operating procedure (SOP). Eur Biophys J. 2021;50(3-4):363-71.

191. Paketurytė V, Petrauskas V, Zubrienė A, Abian O, Bastos M, Chen WY, et al. Uncertainty in protein–ligand binding constants: asymmetric confidence intervals versus standard errors. Eur Biophys J [Internet]. 2021;50(3-4):661-70. Disponible en: https://doi.org/10.1007/s00249-021-01518-4

192. Information P. EnzChek ® Protease Assay Kits. 2004;

193. Papageorgiou AC, Mattsson J. Protein Structure Validation and Analysis with X-Ray Crystallography. Methods Mol Biol. 2014;1129:331-40.

194. Kabsch W. XDS. Acta Crystallogr sect D Biol Crystallogr. 2010;66:125-32.

195. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr Sect D Biol Crystallogr. 2010;66(2):133-44.

196. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010;66(2):213-21.

197. Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr Sect D Biol Crystallogr. 2004;60(12 I):2210-21.

198. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr. 2011;67(4):235-42.

199. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(4):658-74.

200. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: Fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015;43(W1):443-7.

201. Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, et al. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Res. 2021;49:530-4.

202. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. J Mol Biol. 2007;372(3):774-97.

203. DeLano WL. The case for open-source software in drug discovery. Drug Discov Today. 2005;10(3):213-7.

204. Weinstein M. Performance Standards for Antimicrobial Susceptibility Testing. Vol. 8, Journal of Services Marketing. 2021. 18-260 p.

205. García-Sánchez JE, García-Sánchez E, García-García MI. Estudios de sensibilidad en bacterias anaerobias. Enferm Infecc Microbiol Clin. 2014;32(Supl 1):23-9.

206. Bacic MK, Smith CJ. Laboratory maintenance and cultivation of bacteroides species. Curr Protoc Microbiol. 2008;(SUPPL. 9):1-21.

207. Becton D and C. GasPak ™ EZ Gas Generating Container Systems. 2018;8010412(05):21.

208. Veiga A, Toledo M da GT, Rossa LS, Mengarda M, Stofella NCF, Oliveira LJ, et al. Colorimetric microdilution assay: Validation of a standard method for determination of MIC, IC50%, and IC90% of antimicrobial compounds. J Microbiol Methods. 2019;162(May):50-61.

209. Girirajan S, Campbell C, Eichler E. Synergy and antagonism in natural product extracts: when 1 + 1 does not equal 2. Physiol Behav. 2011;176(5):139-48.

210. Jia J, Zhu F, Ma X, Cao ZW, Li YX, Chen YZ. Mechanisms of drug combinations: Interaction and network perspectives. Nat Rev Drug Discov. 2009;8(2):111-28.

211. Gómara M, Ramón-García S. The FICI paradigm: Correcting flaws in antimicrobial in vitro synergy screens at their inception. Biochem Pharmacol. 2019;163:299-307.

212. van Meerloo J, Kaspers GJL, Cloos J. Cell sensitivity assays: the MTT assay. Cancer Cell Cult. 2011;731:237-45.

213. Dunn MJ, Woodruff AL, Anderson MZ. The galleria mellonella waxworm infection model for disseminated candidiasis. J Vis Exp. 2018;2018(141):1-8.

214. Hawe A, Sutter M, Jiskoot W. Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res. 2008;25(7):1487-99.

215. Liebschner D, Afonine P V., Moriarty NW, Poon BK, Sobolev O V., Terwilliger TC, et al. Polder maps: Improving OMIT maps by excluding bulk solvent. Acta Crystallogr Sect D Struct Biol. 2017;73(2):148-57.

216. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr. 2004;60(12 I):2126-32.

217. De Francesco R, Urbani A, Nardi MC, Tomei L, Steinkühler C, Tramontano A. A zinc binding site in viral serine proteinases. Biochemistry. 1996;35(41):13282-7.

218. Wright DL, Holloway JH, Reilley CN. Heats and Entropies of Formation of Metal Chelates of Polyamine and Polyaminocarboxylate Ligands. Anal Chem. 1965;37(7):884-92.

219. Sill??n , Martell, Arthur E., Bjerrum, Jannik,, LG. Stability constants of metal-ion complexes. London: Chemical Society; 1964.

220. Abian O, Vega S, Neira JL, Velazquez-Campoy A. Conformational stability of hepatitis C

virus NS3 protease. Biophys J. 2010;99(11):3811-20.

221. Colvin RA, Holmes WR, Fontaine CP, Maret W. Cytosolic zinc buffering and muffling: Their role in intracellular zinc homeostasis. Metallomics. 2010;2(5):306-17.

222. Outten CE, O'Halloran T V. Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. Science (80-). 2001;292(5526):2488-92.

223. Sensi SL, Paoletti P, Bush AI, Sekler I. Zinc in the physiology and pathology of the CNS. Nat Rev Neurosci. 2009;10(11):780-91.

224. Hara T, Takeda T aki, Takagishi T, Fukue K, Kambe T, Fukada T. Physiological roles of zinc transporters: molecular and genetic importance in zinc homeostasis. J Physiol Sci. 2017;67(2):283-301.

225. Fraga H, Papaleo E, Vega S, Velazquez-Campoy A, Ventura S. Zinc induced folding is essential for TIM15 activity as an mtHsp70 chaperone. Biochim Biophys Acta - Gen Subj. 2013;1830(1):2139-49.

226. Tsai CJ, Nussinov R. A Unified View of «How Allostery Works». PLoS Comput Biol. 2014;10(2):e1003394.

227. Wyman J, Gill SJ. Binding and Linkage: Functional Chemistry of Biological Macromolecules. University Science Books, Mill Valley. 1990.

228. Odds FC. Synergy, antagonism, and what the chequerboard puts between them. J Antimicrob Chemother. 2003;52(1):1.

229. James Booth S. Fusobacterium Infections [Internet]. Reference Module in Biomedical Sciences. Elsevier Inc.; 2014. 1-3 p. Disponible en: http://dx.doi.org/10.1016/B978-0-12-801238-3.05030-3

230. Könönen E, Bryk A, Niemi P, Kanervo-Nordström A. Antimicrobial susceptibilities of Peptostreptococcus anaerobius and the newly described Peptostreptococcus stomatis isolated from various human sources. Antimicrob Agents Chemother. 2007;51(6):2205-7.

231. Bornside GH, Donovan WE, Myers MB. Intracolonic Tensions of Oxygen and Carbon Dioxide in Germfree, Conventional, and Gnotobiotic Rats. Proc Soc Exp Biol Med. 1976;151(2):437-41.

232. Thelaus J, Lundmark E, Lindgren P, Sjödin A, Forsman M. Galleria mellonella reveals

niche differences between highly pathogenic and closely related strains of Francisella spp. Front Cell Infect Microbiol. 2018;8(JUN):1-10.

233. Wagley S, Borne R, Harrison J, Baker-Austin C, Ottaviani D, Leoni F, et al. Galleria mellonella as an infection model to investigate virulence of vibrio parahaemolyticus. Virulence. 2018;9(1):197-207.

234. Kandiah E, Carriel D, Garcia PS, Felix J, Banzhaf M, Kritikos G, et al. Structure, Function, and Evolution of the Pseudomonas aeruginosa Lysine Decarboxylase LdcA. Structure. 2019;27(12):1842-1854.e4.

235. Lange A, Beier S, Huson DH, Parusel R, Iglauer F, Frick JS. Genome sequence of Galleria mellonella (greater wax moth). Genome Announc. 2018;6(2):9-10.

236. Mannala GK, Izar B, Rupp O, Schultze T, Goesmann A, Chakraborty T, et al. Listeria monocytogenes induces a virulence-dependent microRNA signature that regulates the immune response in Galleria mellonella. Front Microbiol. 2017;8(DEC):1-12.

237. Ball P. Quinolone generations: Natural history or natural selection? J Antimicrob Chemother. 2000;46(TOPIC T1):17-24.

238. Emmerson AM, Jones AM. The quinolones: Decades of development and use. J Antimicrob Chemother. 2003;51(SUPPL. 1):13-20.

239. Janknegt R, Centre ZM. Fluorinated quinolones: a review of their mode of action, antimicrobial activity, pharmacokinetics and clinical efficacy. Pharm Weekbl Sci Ed. 1986;8(March 1986).

240. Mevius DJ, Breukink HJ, Van Miert ASJPAM. In vitro activity of flumequine in comparison with several other antimicrobial agents against five pathogens isolated in calves in the Netherlands. Vet Q. 1990;12(4):212-20.

241. DORRESTEIN GM, GOGH H VAN, BUITELAAR MN, NOUWS JFM. Clinical pharmacology and pharmacokinetics of flumequine after intravenous, intramuscular and oral administration in pigeons (Columba livid). J Vet Pharmacol Ther. 1983;6(4):281-92.

242. Rodríguez JM, Diez MJ, Sierra M, Garcia JJ, Fernandez N, Diez R, et al. Distribution of flumequine in intestinal contents and colon tissue in pigs after its therapeutic use in the drinking water. Animals. 2021;11(6):1-11.

243. Villa R, Cagnardi P, Acocella F, Massi P, Anfossi P, Asta F, et al. Pharmacodynamics and

pharmacokinetics of flumequine in pigs after single intravenous and intramuscular administration. Vet J. 2005;170(1):101-7.

244. Elema MO, Hoff KA, Kristensen HG. Bioavailability of flumequine after oral administration to Atlantic salmon (Salmo salar L.). Aquaculture. 1995;136(3-4):209-19.

245. Drlica K, Zhao X. DNA gyrase, topoisomerase IV, and the 4-quinolones. Microbiol Mol Biol Rev. 1997;61(3):377-92.

246. Nitiss JL. DNA topoisomerase II and its growing repertoire of biological functions. Nat Rev Cancer. 2009;9(5):327-37.

247. Gellert M, Mizuuchi K, O'Dea MH, Nash HA. DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc Natl Acad Sci U S A. 1976;73(11):3872-6.

248. Sociedad Española de Oncología Médica (SEOM), Red Española de Registros del Cáncer (REDECAN). Las cifras del cáncer en España.2022. 2022. 2022. 1-38 p.

249. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71(1):7-33.

250. Park W, Chawla A, Reilly EMO. Pancreatic Cancer: A Review. 2022;326(9):851-62.

251. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states. Cancer Res. 2014;74(11):2913-21.

252. Conroy T, Desseigne F, Ychou M, Bouché O, Guimbaud R, Bécouarn Y, et al. FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. N Engl J Med. 2011;364(19):1817-25.

253. Conroy T, Hammel P, Hebbar M, Ben Abdelghani M, Wei AC, Raoul J-L, et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N Engl J Med. 2018;379(25):2395-406.

254. Hoff DD Von, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med. 2013;369(18):1691-703.

255. Bosetti C, Lucenteforte E, Silverman DT, Petersen G, Bracci PM, Ji BT, et al. Cigarette smoking and pancreatic cancer: An analysis from the International Pancreatic Cancer Case-Control Consortium (PANC4). Ann Oncol. 2012;23(7):1880-8.

256. Wang YT, Gou YW, Jin WW, Xiao M, Fang HY. Association between alcohol intake and the risk of pancreatic cancer: A dose-response meta-analysis of cohort studies. BMC Cancer. 2016;16(1):1-11.

257. Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. Lancet Public Heal. 2019;4(3):e137-47.

258. Genkinger JM, Spiegelman D, Anderson KE, Bergkvist L, Bernstein L, Van Den Brandt PA, et al. Alcohol intake and pancreatic cancer risk: A pooled analysis of fourteen cohort studies. Cancer Epidemiol Biomarkers Prev. 2009;18(3):765-76.

259. Golan T, Kindler HL, Park JO, Reni M, Macarulla T, Hammel P, et al. Geographic and ethnic heterogeneity of germline BRCA1 or BRCA2 mutation prevalence among patients with metastatic pancreatic cancer screened for entry into the POLO trial. J Clin Oncol. 2020;38(13):1442-54.

260. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol. 2017;35(30):3382-90.

261. Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. JAMA - J Am Med Assoc. 2018;319(23):2401-9.

262. Sosa V, Moliné T, Somoza R, Paciucci R, Kondoh H, LLeonart ME. Oxidative stress and cancer: An overview. Ageing Res Rev. 2013;12(1):376-90.

263. Schieber M, Chandel NS. ROS function in redox signaling. Curr Biol. 2014;24(10):453-62.

264. Martinez-Useros J, Li W, Cabeza-Morales M, Garcia-Foncillas J. Oxidative stress: A new target for pancreatic cancer prognosis and treatment. J Clin Med. 2017;6(3):1-16.

265. Terhune PG, Phifer DM, Tosteson TD, Longnecker DS. K-ras mutation in focal proliferative lesions of human pancreas. Cancer Epidemiol Biomarkers Prev. 1998;7(6):515-21.

266. Manuscript A, Neoplasia I. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology. 2012;142(4):730-3.

267. Murtaugh L. C. Pathogenesis of pancreatic cancer: lessons from animal models. Toxicol

Pathol. 2014;42(1):217-28.

268. Walter FM, Mills K, Mendonça SC, Abel GA, Basu B, Carroll N, et al. Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. Lancet Gastroenterol Hepatol. 2016;1(4):298-306.

269. Wang S, Zheng Y, Yang F, Zhu L, Zhu XQ, Wang ZF, et al. The molecular biology of pancreatic adenocarcinoma: translational challenges and clinical perspectives. Signal Transduct Target Ther. 2021;6(1).

270. Werner J, Combs SE, Springfeld C, Hartwig W, Hackert T, Büchler MW. Advanced-stage pancreatic cancer: Therapy options. Nat Rev Clin Oncol. 2013;10(6):323-33.

271. Vienot A, Chevalier H, Bolognini C, Gherga E, Klajer E, Meurisse A, et al. FOLFOXIRI vs FOLFIRINOX as first-line chemotherapy in patients with advanced pancreatic cancer: A population-based cohort study. World J Gastrointest Oncol. 2020;12(3):332-46.

272. Riedl JM, Posch F, Horvath L, Gantschnigg A, Renneberg F, Schwarzenbacher E, et al. Gemcitabine/nab-Paclitaxel versus FOLFIRINOX for palliative first-line treatment of advanced pancreatic cancer: A propensity score analysis. Eur J Cancer. 2021;151:3-13.

273. Wang-gillam A, Li C, Bodoky G, Dean A, Shan Y, Jameson G, et al. Nanoliposomal irinotecan with fluorouracil and folinic acid in metastatic pancreatic cancer after previous gemcitabine-based therapy (NAPOLI-1): a global, randomised, open-label, phase 3 trial. 2015;6736(15).

274. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16(1):18-29.

275. Dunker KA, Lawson DJ, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically Disordered Protein. J Mol Graph Model. 2001;19(00):26-59.

276. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. Crystal structure of the lactose operon represser and its complexes with DNA and inducer. Science (80-). 1996;271(5253):1247-54.

277. Spolar RS, Record MT. Coupling of local folding to site-specific binding of proteins to DNA. Science (80-). 1994;263(5148):777-84.

278. Aviles F, Chapman G, Kneale G, Crane-Robinson C, Bradbur y E. The Conformation of

Histone H5. Eur J Biochem. 1978;371(88):363-71.

279. Wright PE, Dyson HJ. Intrinsically Unstructured Proteins : Re-assessing the Protein Structure-Function Paradigm. J Mol Biol. 1999;293:321-31.

280. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. Annu Rev Biophys. 2008;37:215-46.

281. Plaxco KW, Grob M. The importance of being unfolded.

282. Neira JL, Rizzuti B, Jiménez-Alesanco A, Palomino-Schätzlein M, Abián O, Velázquez-Campoy A, et al. A phosphorylation-induced switch in the nuclear localization sequence of the intrinsically disordered nupr1 hampers binding to importin. Biomolecules. 2020;10(9):1-22.

283. Bah A, Forman-Kay JD. Modulation of intrinsically disordered protein function by post-translational modifications. J Biol Chem. 2016;291(13):6696-705.

284. Cano CE, Iovanna JL. Stress proteins and pancreatic cancer metastasis. ScientificWorldJournal. 2010;10:1958-66.

285. Calderwood SK, Khaleque MA, Sawyer DB, Ciocca DR. Heat shock proteins in cancer: Chaperones of tumorigenesis. Trends Biochem Sci. 2006;31(3):164-72.

286. Chowdhury UR, Samant RS, Fodstad O, Shevde LA. Emerging role of nuclear protein 1 (NUPR1) in cancer biology. Cancer Metastasis Rev. 2009;28(1-2):225-32.

287. Mallo GV, Fiedler F, Calvo EL, Ortiz EM, Vasseur S, Keim V, et al. Cloning and expression of the rat p8 cDNA, a new gene activated in pancreas during the acute phase of pancreatitis, pancreatic development, and regeneration, and which promotes cellular growth. J Biol Chem. 1997;272(51):32360-9.

288. Valacco MP, Varone C, Malicet C, Cánepa E, Iovanna JL, Moreno S. Cell growth-dependent subcellular localization of p8. J Cell Biochem. 2006;97(5):1066-79.

289. Cano CE, Hamidi T, Garcia MN, Grasso D, Loncle C, Garcia S, et al. Genetic inactivation of Nupr1 acts as a dominant suppressor event in a two-hit model of pancreatic carcinogenesis. Gut. 2014;63(6):984-95.

290. Hamidi T, Algül H, Cano CE, Sandi MJ, Molejon MI, Riemann M, et al. Nuclear protein 1 promotes pancreatic cancer development and protects cells from stress by inhibiting apoptosis. J Clin Invest. 2012;122(6):2092-103.

291. Martin TA, Li AX, Sanders AJ, Ye L, Frewer K, Hargest R, et al. NUPR1 and its potential role in cancer and pathological conditions (Review). Int J Oncol. 2021;58(5):1-12.

292. Lan W, Santofimia-Castaño P, Swayden M, Xia Y, Zhou Z, Audebert S, et al. ZZW-115-dependent inhibition of NUPR1 nuclear translocation sensitizes cancer cells to genotoxic agents. JCI Insight. 2020;5(18):1-19.

293. Sandi MJ, Hamidi T, Malicet C, Cano C, Loncle C, Pierres A, et al. p8 Expression controls pancreatic cancer cell migration, invasion, adhesion, and tumorigenesis. J Cell Physiol. 2011;226(12):3442-51.

294. Goruppi S, Iovanna JL. Stress-inducible protein p8 is involved in several physiological and pathological processes. J Biol Chem. 2010;285(3):1577-81.

295. Goruppi S, Kyriakis JM. The pro-hypertrophic basic helix-loop-helix protein p8 is degraded by the ubiquitin/proteasome system in a protein kinase B/Akt- and glycogen synthase kinase-3-dependent manner, whereas endothelin induction of p8 mRNA and renal mesangial cell hypertrophy. J Biol Chem. 2004;279(20):20950-8.

296. Santofimia-Castaño P, Rizzuti B, Abián O, Velázquez-Campoy A, Iovanna JL, Neira JL. Amphipathic helical peptides hamper protein-protein interactions of the intrinsically disordered chromatin nuclear protein 1 (NUPR1). Biochim Biophys Acta - Gen Subj. 2018;1862(6):1283-95.

297. Santofimia-Castaño P, Rizzuti B, Pey ÁL, Soubeyran P, Vidal M, Urrutia R, et al. Intrinsically disordered chromatin protein NUPR1 binds to the C-terminal region of polycomb RING1B. Proc Natl Acad Sci U S A. 2017;114(31):E6332-41.

298. Aguado-Llera D, Hamidi T, Doménech R, Pantoja-Uceda D, Gironella M, Santoro J, et al. Deciphering the binding between Nupr1 and MSL1 and their DNA-repairing activity. PLoS One. 2013;8(10).

299. Santofimia-Castaño P, Rizzuti B, Xia Y, Abian O, Peng L, Velázquez-Campoy A, et al. Targeting intrinsically disordered proteins involved in cancer. Cell Mol Life Sci. 2020;77(9):1695-707.

300. Pommier RM, Gout J, Vincent DF, Cano CE, Kaniewski B, Martel S, et al. The human NUPR1/P8 gene is transcriptionally activated by transforming growth factor β via the SMAD signalling pathway. Biochem J. 2012;445(2):285-93.

301. Hamidi T, Cano CE, Grasso D, Garcia MN, Sandi MJ, Calvo EL, et al. Nupr1-aurora kinase

a pathway provides protection against metabolic stress-mediated autophagic-associated cell death. Clin Cancer Res. 2012;18(19):5234-46.

302. Gironella M, Malicet C, Cano C, Sandi MJ, Hamidi T, Tauil RMN, et al. p8/nupr1 regulates DNA-repair activity after double-strand gamma irradiation-induced DNA damage. J Cell Physiol. 2009;221(3):594-602.

303. Carracedo A, Lorente M, Egia A, Blázquez C, García S, Giroux V, et al. The stress-regulated protein p8 mediates cannabinoid-induced apoptosis of tumor cells. Cancer Cell. 2006;9(4):301-12.

304. Emma MR, Iovanna JL, Bachvarov D, Puleio R, Loria GR, Augello G, et al. NUPR1, a new target in liver cancer: Implication in controlling cell growth, migration, invasion and sorafenib resistance. Cell Death Dis. 2016;7(6):e2269-14.

305. Santofimia-Castaño P, Lan W, Bintz J, Gayet O, Carrier A, Lomberk G, et al. Inactivation of NUPR1 promotes cell death by coupling ER-stress responses with necrosis. Sci Rep. 2018;8(1):1-16.

306. Brannon KM, Million Passe CM, White CR, Bade NA, King MW, Quirk CC. Expression of the high mobility group A family member p8 is essential to maintaining tumorigenic potential by promoting cell cycle dysregulation in LβT2 cells. Cancer Lett. 2007;254(1):146-55.

307. Kong DK, Georgescu SP, Cano C, Aronovitz MJ, Iovanna JL, Patten RD, et al. Deficiency of the transcriptional regulator p8 results in increased autophagy and apoptosis, and causes impaired heart function. Mol Biol Cell. 2010;21:1335-49.

308. Malicet C, Giroux V, Vasseur S, Dagorn JC, Neira JL, Iovanna JL. Regulation of apoptosis by the p8/prothymosin α complex. Proc Natl Acad Sci U S A. 2006;103(8):2671-6.

309. Grasso D, Garcia MN, Hamidi T, Cano C, Calvo E, Lomberk G, et al. Genetic inactivation of the pancreatitis-inducible gene Nupr1 impairs PanIN formation by modulating KrasG12D-induced senescence. Cell Death Differ. 2014;21(10):1633-41.

310. Grasso D, Bintz J, Lomberk G, Molejon MI, Loncle C, Garcia MN, et al. Pivotal Role of the Chromatin Protein Nupr1 in Kras-Induced Senescence and Transformation. Sci Rep. 2015;5:1-11.

311. Cai D, Huang E, Luo B, Yang Y, Zhang F, Liu C, et al. Nupr1/Chop signal axis is involved in mitochondrion-related endothelial cell apoptosis induced by methamphetamine. Cell

Death Dis. 2016;7(3):1-14.

312. González-Arzola K, Guerra-Castellano A, Rivero-Rodríguez F, Casado-Combreras M, Pérez-Mejías G, Díaz-Quintana A, et al. Mitochondrial cytochrome c shot towards histone chaperone condensates in the nucleus. FEBS Open Bio. 2021;11(9):2418-40.

313. González-Arzola K, Velázquez-Cruz A, Guerra-Castellano A, Casado-Combreras M, Pérez-Mejías G, Díaz-Quintana A, et al. New moonlighting functions of mitochondrial cytochrome c in the cytoplasm and nucleus. FEBS Lett. 2019;593(22):3101-19.

314. Shiraki M, Xu X, Iovanna JL, Kukita T, Hirata H, Kamohara A, et al. Deficiency of stress-associated gene Nupr1 increases bone volume by attenuating differentiation of osteoclasts and enhancing differentiation of osteoblasts. FASEB J. 2019;33(8):8836-52.

315. Hoffmeister A, Ropolo A, Vasseur S, Mallo G V., Bodeker H, Ritz-Laser B, et al. The HMG-I/Y-related protein p8 binds to p300 and Pax2 trans-activation domain-interacting protein to regulate the trans-activation activity of the Pax2A and Pax2B transcription factors on the glucagon gene promoter. J Biol Chem. 2002;277(25):22314-9.

316. Päth G, Opel A, Knoll A, Seufert J. Nuclear Protein p8 Is Associated with Glucose-Induced Pancreatic β-Cell Growth. Diabetes. 2004;53(SUPPL. 1).

317. Maida A, Zota A, Sjøberg KA, Schumacher J, Sijmonsma TP, Pfenninger A, et al. A liver stress-endocrine nexus promotes metabolic integrity during dietary protein dilution. J Clin Invest. 2016;126(9):3263-78.

318. Urrutia R, Velez G, Lin M, Lomberk G, Neira JL, Iovanna J. Evidence supporting the existence of a NUPR1-like family of helix-loop-helix chromatin proteins related to, yet distinct from, AT hook-containing HMG proteins. J Mol Model. 2014;20(8).

319. Neira JL, López MB, Sevilla P, Rizzuti B, Cámara-Artigas A, Vidal M, et al. The chromatin nuclear protein NUPR1L is intrinsically disordered and binds to the same proteins as its paralogue. Biochem J. 2018;475(14):2271-91.

320. Lopez MB, Garcia MN, Grasso D, Bintz J, Molejon MI, Velez G, et al. Functional Characterization of Nupr1L, A Novel p53-Regulated Isoform of the High-Mobility Group (HMG)-Related Protumoral Protein Nupr1. J Cell Physiol. 2015;230(12):2936-50.

321. Stewart M. Molecular mechanism of the nuclear protein import cycle. Nat Rev Mol Cell Biol. 2007;8(3):195-208.

322. Bednenko J, Cingolani G, Gerace L. Nucleo-cytoplasmic transport: Navigating the channel. Traffic. 2003;4(3):127-35.

323. Oka M, Yoneda Y. Importin α: Functions as a nuclear transport factor and beyond. Proc Japan Acad Ser B Phys Biol Sci. 2018;94(7):259-74.

324. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA. Importin α: A multipurpose nuclear-transport receptor. Trends Cell Biol. 2004;14(9):505-14.

325. Miyamoto Y, Loveland KL, Yoneda Y. Nuclear importin α and its physiological importance. Commun Integr Biol. 2012;5(2):220-2.

326. Mason DA, Stage DE, Goldfarb DS. Evolution of the metazoan-specific importin α gene family. J Mol Evol. 2009;68(4):351-65.

327. Pumroy RA, Ke S, Hart DJ, Zachariae U, Cingolani G. Molecular determinants for nuclear import of influenza A PB2 by importin α isoforms 3 and 7. Structure. 2015;23(2):374-84.

328. Kobe B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. Nat Struct Biol. 1999;6(4):388-97.

329. Top 10 Emerging Technologies 2019. World Econ Forum. 2019;(June):4-15.

330. Marques L de O, Soares B, Silva de Lima M. Trifluoperazine for schizophrenia. Cochrane Database Syst Rev. 2004;

331. Santofimia-Castaño P, Xia Y, Lan W, Zhou Z, Huang C, Peng L, et al. Ligand-based design identifies a potent NUPR1 inhibitor exerting anticancer activity via necroptosis. J Clin Invest. 2019;129(6):2500-13.

332. Gintant G, Sager PT, Stockbridge N. Evolution of strategies to improve preclinical cardiac safety testing. Nat Rev Drug Discov. 2016;15(7):457-71.

333. Fermini B, Fossa AA. The impact of drug-induced QT interval prolongation on drug discovery and development. Nat Rev Drug Discov. 2003;2(6):439-47.

334. Garrido A, Lepailleur A, Mignani SM, Dallemagne P, Rochais C. hERG toxicity assessment : useful guidelines for drug design. Eur J Med Chem. 2020;195(112290).

335. Kalyaanamoorthy S, Barakat KH. Development of Safe Drugs: The hERG Challenge. Med Res Rev. 2017;(0):1-31.

336. Neira JL, Rizzuti B, Jiménez-Alesanco A, Abián O, Velázquez-Campoy A, Iovanna JL. The

paralogue of the intrinsically disordered nuclear protein 1 has a nuclear localization sequence that binds to human importin α3. Int J Mol Sci. 2020;21(19):1-19.

337. Encinar JA, Mallo G V., Mizyrycki C, Giono L, González-Ros JM, Rico M, et al. Human p8 Is a HMG-I/Y-like Protein with DNA Binding Activity Enhanced by Phosphorylation. J Biol Chem. 2001;276(4):2742-51.

338. Neira JL, Rizzuti B, Iovanna JL. Determinants of the pKa values of ionizable residues in an intrinsically disordered protein. Arch Biochem Biophys. 2016;598:18-27.

339. Haverkamp W, Camm AJ, Janse MJ, Rosen MR, Antzelevitch C, Escande D, et al. The potential for QT prolongation and pro-arrhythmia by non-anti-arrhythmic drugs : Clinical and regulatory implications Report on a Policy Conference of the European Society of Cardiology q. Cardiovasc Res. 2000;47:219-33.

340. Lester RM, Paglialunga S, Johnson IA. QT Assessment in Early Drug Development : The Long and the Short of It. Int J Mol Sci. 2019;20(1324).

341. Waterhouse NJ. The cellular energy crisis: Mitochondria and cell death. Med Sci Sports Exerc. 2003;35(1):105-10.

342. Cano CE, Hamidi T, Sandi MJ, Iovanna JL. Nupr1: The Swiss-knife of cancer. Journal of Cellular Physiology. 2011;1439-43.

343. Santofimia-Castaño P, Xia Y, Peng L, Velázquez-Campoy A, Abián O, Lan W, et al. Targeting the Stress-Induced Protein NUPR1 to Treat Pancreatic Adenocarcinoma. Cells. 2019;8(11):1-9.

344. Santofimia-Castaño P, Rizzuti B, Xia Y, Abian O, Peng L, Velázquez-Campoy A, et al. Designing and repurposing drugs to target intrinsically disordered proteins for cancer treatment: using NUPR1 as a paradigm. Mol Cell Oncol. 2019;6(5):5-7.

345. Garcia-Montero A, Vasseur S, Mallo GV, Soubeyran P, Dagorn JC, Iovanna JL. Expression of the stress-induced p8 mRNA is transiently activated after culture medium change. Eur J Cell Biol. 2001;80(11):720-5.

346. Chen CY, Wu SM, Lin YH, Chi HC, Lin SL, Yeh CT, et al. Induction of nuclear protein-1 by thyroid hormone enhances platelet-derived growth factor a mediated angiogenesis in liver cancer. Theranostics. 2019;9(8):2361-79.

347. Kim K-S, Jin D-I, Yoon S, Baek S-Y, Kim B-S, Oh S-O. Expression and roles of NUPR1 in

cholangiocarcinoma cells. Anat Cell Biol. 2012;45(1):17.

348. Guo X, Wang W, Hu J, Feng K, Pan Y, Zhang L, et al. Lentivirus-Mediated RNAi Knockdown of NUPR1 Inhibits Human Nonsmall Cell Lung Cancer Growth In Vitro and In Vivo. Anat Rec. 2012;295(12):2114-21.

349. Li Y, Yin Y, Ma J, Sun Y, Zhou R, Cui B, et al. Combination of AAV-mediated NUPR1 knockdown and trifluoperazine induces premature senescence in human lung adenocarcinoma A549 cells in nude mice. Oncol Rep. 2020;43(2):681-8.

350. Zhou C, Xu J, Lin J, Lin R, Chen K, Kong J, et al. Long noncoding RNA FEZF1-AS1 promotes osteosarcoma progression by regulating the miR-4443/NUPR1 axis. Oncol Res. 2018;26(9):1335-43.

351. Zeng C, Li X, Li A, Yi B, Peng X, Huang X, et al. Knockdown of NUPR1 inhibits the growth of U266 and RPMI8226 multiple myeloma cell lines via activating PTEN and caspase activation-dependent apoptosis. Oncol Rep. 2018;40(3):1487-94.

352. Li A, Li X, Chen X, Zeng C, Wang Z, Li Z, et al. NUPR1 Silencing Induces Autophagy-Mediated Apoptosis in Multiple Myeloma Cells through the PI3K/AKT/mTOR Pathway. DNA Cell Biol. 2020;39(3):368-78.

353. Li J, Ren S, Liu Y, Lian Z, Dong B, Yao Y, et al. Knockdown of NUPR1 inhibits the proliferation of glioblastoma cells via ERK1/2, p38 MAPK and caspase-3. J Neurooncol. 2017;132(1):15-26.

354. Yu J, Zhu H, Li R, Jiang Q, Luan W, Shi J, et al. Oncogenic role of NUPR1 in ovarian cancer. Onco Targets Ther. 2020;13:12289-300.

355. Miyatake H, Sanjoh A, Unzai S, Matsuda G, Tatsumi Y, Miyamoto Y, et al. Crystal structure of human importin-α1 (Rch1), revealing a potential autoinhibition mode involving homodimerization. PLoS One. 2015;10(2):1-14.

356. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene. 1989;77(1):51-9.

357. Higuchi R, Krummel B, Saiki R. A general method of in vitro preparation and specific mutagenesis of dna fragments: Study of protein and DNA interactions. Nucleic Acids Res. 1988;16(15):7351-67.

358. Marvaldi L, Panayotis N, Alber S, Dagan SY, Okladnikov N, Koppel I, et al. Importina α3

regulates chronic pain pathways in peripheral sensory neurons. 2020;369(6505):842-6.

359. Smith KM, Tsimbalyuk S, Edwards MR, Cross EM, Batra J, Soares da Costa TP, et al. Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses. Nat Commun. 2018;9(1):1-13.

360. Cingolani G, Bednenko J, Gillespie MT, Gerace L. Molecular basis for the recognition of a nonclassical nuclear localization signal by importin β. Mol Cell. 2002;10(6):1345-53.

361. Sankhala RS, Lokareddy RK, Begum S, Pumroy RA, Gillilan RE, Cingolani G. Three-dimensional context rather than NLS amino acid sequence determines importin α subtype specificity for RCC1. Nat Commun. 2017;8(1):1-14.

362. Junod SL, Kelich JM, Ma J, Yang W. Nucleocytoplasmic transport of intrinsically disordered proteins studied by high-speed super-resolution microscopy. Protein Sci. 2020;29(6):1459-72.

363. Neira JL, Correa J, Rizzuti B, Santofimia-Castaño P, Abian O, Velázquez-Campoy A, et al. Dendrimers as Competitors of Protein-Protein Interactions of the Intrinsically Disordered Nuclear Chromatin Protein NUPR1. Biomacromolecules. 2019;20(7):2567-76.

364. Yadahalli S, Neira JL, Johnson CM, Tan YS, Pamela JER, Chattopadhy A, et al. Kinetic and thermodynamic effects of phosphorylation on p53 binding to MDM2. 2019;(July 2018):1-15.

365. Muhandiram R, Zhao C, Forman-kay JD, Sonenberg N, Kay LE. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. 2014;

366. Gandhi NS, Landrieu I, Byrne C, Kukic P, Amniai L, Fran Å, et al. A Phosphorylation-Induced Turn Defines the Alzheimer s Disease AT8 Antibody Epitope on the Tau Protein. 2015;6819-23.

367. Kumar A, Narayanan V, Sekhar A. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. Biochemistry. 2020;59(1):57-73.

368. Das RK, Pappu R V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110(33):13392-7.

369. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu R V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A. 2010;107(18):8183-8.

370. Gibbs EB, Lu F, Portz B, Fisher MJ, Medellin BP, Laremore TN, et al. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. Nat Commun. 2017;8(May):1-11.

371. Xiang S, Gapsys V, Kim HY, Bessonov S, Hsiao HH, Möhlmann S, et al. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. Structure. 2013;21(12):2162-74.

372. Turner AL, Watson M, Wilkins OG, Cato L, Travers A, Thomas JO, et al. Highly disordered histone H1−DNA model complexes and their condensates. Proc Natl Acad Sci U S A. 2018;115(47):11964-9.

373. Banavali NK, Roux B. Anatomy of a Structural Pathway for Activation of the Catalytic Domain of Src Kinase Hck. Proteins: Struct Funct Bioinform. 2007;67(2007):1096–1112.

374. Hendus-Altenburger R, Lambrughi M, Terkelsen T, Pedersen SF, Papaleo E, Lindorff-Larsen K, et al. A phosphorylation-motif for tuneable helix stabilisation in intrinsically disordered proteins – Lessons from the sodium proton exchanger 1 (NHE1). Cell Signal. 2017;37(January):40-51.

375. Espinoza-Fonseca LM, Kast D, Thomas DD. Molecular dynamics simulations reveal a disorder-to-order transition on phosphorylation of smooth muscle myosin. Biophys J [Internet]. 2007;93(6):2083-90. Disponible en: http://dx.doi.org/10.1529/biophysj.106.095802

376. Chu IM, Hengst L, Slingerland JM. The Cdk inhibitor p27 in human cancer: Prognostic potential and relevance to anticancer therapy. Nat Rev Cancer. 2008;8(4):253-67.

377. He Y, Chen Y, Mooney SM, Rajagopalan K, Bhargava A, Sacho E, et al. Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen. J Biol Chem. 2015;290(41):25090-102.

378. Gilbert RJC, Gordiyenko Y, Von Der Haar T, Sonnen AFP, Hofmann G, Nardelli M, et al. Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex. Proc Natl Acad Sci U S A. 2007;104(14):5788-93.

379. Boulikas T. Putative nuclear localization signals (NLS) in protein transcription factors. J

Cell Biochem. 1994;55(1):32-58.

380. Boisvert M, Bouchard-Lévesque V, Fernandes S, Tijssen P. Classic Nuclear Localization Signals and a Novel Nuclear Localization Motif Are Required for Nuclear Transport of Porcine Parvovirus Capsid Proteins. J Virol. 2014;88(20):11748-59.

381. Hill R, Cautain B, De Pedro N, Link W. Targeting nucleocytoplasmic transport in cancer therapy. Oncotarget. 2014;5(1):11-28.

382. Jans DA, Martin AJ, Wagstaff KM. Inhibitors of nuclear transport. Curr Opin Cell Biol. 2019;58:50-60.

383. Kosyna FK, Depping R. Controlling the gatekeeper: Therapeutic targeting of nuclear transport. Cells. 2018;7(11).

384. Weber DJ, Markowitz J, Carrier F, Spring S, Mackerell AD. Inhibitors of the S100-p53 protein-protein interaction and method of inhibiting cancer employing the same. United States; US 8,053,477 B2, 2011.

385. Joseph Markowitz, David J. Weber. A Search for Inhibitors of S100B, a Member of the S100 Family of Calcium-Binding Proteins. Mini-Reviews Med Chem. 2007;7(6):609-16.

386. Lin J, Yang Q, Yan Z, Markowitz J, Wilder PT, Carrier F, et al. Inhibiting S100B restores p53 levels in primary malignant melanoma cancer cells. J Biol Chem [Internet]. 2004;279(32):34071-7. Disponible en: http://dx.doi.org/10.1074/jbc.M405419200

387. Markowitz J, Rustandi RR, Varney KM, Wilder PT, Udan R, Su LW, et al. Calcium-binding properties of wild-type and EF-hand mutants of S100B in the presence and absence of a peptide derived from the C-terminal negative regulatory domain of p53. Biochemistry. 2005;44(19):7305-14.

388. Ka-Ming Chan F, Moriwaki K, De Rosa MJ. Detection of Necrosis by Release of Lactate Dehydrogenase (LDH) Activity. Methods Mol Biol. 2013;979:65-70.

389. Challa S, Ka-Ming Chan F. Going up in Flames: Necrotic Cell Injury and Inflammatory Diseases. Cell Mol Life Sci. 2010;67(19):3241–3253.

390. Moquin D, Ka-Ming Chan F. The Molecular Regulation of Programmed Necrotic Cell Injury. Trends Biochem Sci. 2010;35(8):434–441.

391. Chan FKM, Shisler J, Bixby JG, Felices M, Zheng L, Appel M, et al. A Role for Tumor Necrosis Factor Receptor-2 and Receptor-interacting Protein in Programmed Necrosis

and Antiviral Responses. J Biol Chem [Internet]. 2003;278(51):51613-21. Disponible en: http://dx.doi.org/10.1074/jbc.M305633200

392. Walsh JG, Cullen SP, Sheridan C, Lüthi AU, Gerner C, Martin SJ. Executioner caspase-3 and caspase-7 are functionally distinct proteases. Proc Natl Acad Sci U S A. 2008;105(35):12815-9.

393. Elmore S. Apoptosis: A Review of Programmed Cell Death. Toxicol Pathol. 2007;35(4):495-516.

394. Jonathan R. Friedman and Jodi Nunnari. Mitochondrial form and function Jonathan. Nano [Internet]. 2008;6(9):2166-71. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4075653/

395. Zhang J, Han X, Lin Y. Dissecting the regulation and function of ATP at the single-cell level. PLoS Biol. 2018;16(12):1-7.

396. Long JE, Jankovic M, Maddalo D. Drug discovery oncology in a mouse: Concepts, models and limitations. Futur Sci OA. 2021;7(8).

397. Rygaard J, Povlsen CO. Heterotransplantation of a human malignant tumour to «Nude» mice. Acta path microbiol scand. 1969;77:758–760.

398. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84.

399. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov Today [Internet]. 2021;26(4):1040-52. Disponible en: https://doi.org/10.1016/j.drudis.2020.11.037

400. Subbiah V. The next generation of evidence-based medicine. Nat Med. 2023;29(1):49-58.

401. Gaddy JA, Actis LA, Arivett BA, Mcconnell MJ, Rafael LR, Pachón J. Role of Acinetobactin-mediated iron acquisition functions in the interaction of Acinetobacter baumannii strain ATCC 19606T with human lung epithelial cells, Galleria mellonella caterpillars, and mice. Infect Immun. 2012;80(3):1015-24.

402. Peleg AY, Jara S, Monga D, Eliopoulos GM, Moellering RC, Mylonakis E. Galleria mellonella as a model system to study Acinetobacter baumannii pathogenesis and

therapeutics. Antimicrob Agents Chemother. 2009;53(6):2605-9.

403. Chusri S, Chongsuvivatwong V, Silpapojakul K, Singkhamanan K, Hortiwakul T, Charernmak B, et al. Clinical characteristics and outcomes of community and hospital-acquired Acinetobacter baumannii bacteremia. J Microbiol Immunol Infect. 2019;52(5):796-806.

404. Niu T, Guo L, Luo Q, Zhou K, Yu W, Chen Y, et al. Wza gene knockout decreases Acinetobacter baumannii virulence and affects Wzy-dependent capsular polysaccharide synthesis. Virulence. 2020;11(1):1-13.

405. Aperis G, Burgwyn Fuchs B, Anderson CA, Warner JE, Calderwood SB, Mylonakis E. Galleria mellonella as a model host to study infection by the Francisella tularensis live vaccine strain. Microbes Infect. 2007;9(6):729-34.

406. Jander G, Rahme LG, Ausubel FM. Positive correlation between virulence of Pseudomonas aeruginosa mutants in mice and insects. J Bacteriol. 2000;182(13):3843-5.

407. Miyata S, Casey M, Frank DW, Ausubel FM, Drenkard E. Use of the Galleria mellonella caterpillar as a model host to study the role of the type III secretion system in Pseudomonas aeruginosa pathogenesis. Infect Immun. 2003;71(5):2404-13.

408. Medina-Rojas M, Stribling W, Snesrud E, Garry BI, Li Y, Gann PM, et al. Comparison of Pseudomonas aeruginosa strains reveals that Exolysin A toxin plays an additive role in virulence. Pathog Dis. 2020;78(1):1-11.

409. Kaszab E, Radó J, Kriszt B, Pászti J, Lesinszki V, Szabó A, et al. Groundwater, soil and compost, as possible sources of virulent and antibiotic-resistant Pseudomonas aeruginosa. Int J Environ Health Res. 2021;31(7):848-60.

410. Champion OL, Cooper IAM, James SL, Ford D, Karlyshev A, Wren BW, et al. Galleria mellonella as an alternative infection model for Yersinia pseudotuberculosis. Microbiology. 2009;155(5):1516-22.

411. Desbois AP, Coote PJ. Wax moth larva (Galleria mellonella): An in vivo model for assessing the efficacy of antistaphylococcal agents. J Antimicrob Chemother. 2011;66(8):1785-90.

412. Li L, Chen H, Liu Y, Xu S, Wu M, Liu Z, et al. Synergistic effect of linezolid with fosfomycin against Staphylococcus aureus in vitro and in an experimental Galleria mellonella

model. J Microbiol Immunol Infect. 2020;53(5):731-8.

413. Forti F, Roach DR, Cafora M, Pasini ME, Horner DS, Fiscarelli E V., et al. Design of a broad-range bacteriophage cocktail that reduces pseudomonas aeruginosa biofilms and treats acute infections in two animal models. Antimicrob Agents Chemother. 2018;62(6):1-13.

414. Olsen RJ, Ebru Watkins M, Cantu CC, Beres SB, Musser JM. Virulence of serotype M3 group a streptococcus strains in wax worms (Galleria mellonella larvae). Virulence. 2011;2(2):111-9.

415. Abranches J, Miller JH, Martinez AR, Simpson-Haidaris PJ, Burne RA, Lemos JA. The collagen-binding protein Cnm is required for Streptococcus mutans adherence to and intracellular invasion of human coronary artery endothelial cells. Infect Immun. 2011;79(6):2277-84.

416. Michaux C, Sanguinetti M, Reffuveille F, Auffray Y, Posteraro B, Gilmore MS, et al. SlyA is a transcriptional regulator involved in the virulence of Enterococcus faecalis. Infect Immun. 2011;79(7):2638-45.

417. Yasmin A, Kenny JG, Shankar J, Darby AC, Hall N, Edwards C, et al. Comparative genomics and transduction potential of Enterococcus faecalis temperate bacteriophages. J Bacteriol. 2010;192(4):1122-30.

418. Fuchs BB, Eby J, Nobile CJ, El Khoury JB, Mitchell AP, Mylonakis E. Role of filamentation in Galleria mellonella killing by Candida albicans. Microbes Infect. 2010;12(6):488-96.

419. Mylonakis E, Moreno R, Khoury JBE, Idnurm A, Heitman J, Calderwood SB, et al. Galleria mellonella as a model system to study Cryptococcus neoformans pathogenesis. Infect Immun. 2005;73(7):3842–3850.

420. Brochado AR, Telzerow A, Bobonis J, Banzhaf M, Mateus A, Selkrig J, et al. Species-specific activity of antibacterial drug combinations. Nature. 2018;559(7713):259-63.

421. Manohar P, Nachimuthu R, Lopes BS. The therapeutic potential of bacteriophages targeting gram-negative bacteria using Galleria mellonella infection model. BMC Microbiol. 2018;18(1):1-11.

422. De la Cruz MA, Morgan JK, Ares MA, Yáñez-Santos JA, Riordan JT, Girón JA. The two-component system CpxRA negatively regulates the locus of enterocyte effacement of enterohemorrhagic Escherichia coli involving σ32 and lon protease. Front Cell Infect

Microbiol. 2016;6(FEB).

423. Mills G, Dumigan A, Kidd T, Hobley L, Bengoechea JA. Identification and characterization of two Klebsiella pneumoniae lpxL lipid A late acyltransferases and their role in virulence. Infect Immun. 2017;85(9):1-19.

424. Insua JL, Llobet E, Moranta D, Pérez-Gutiérrez C, Tomás A, Garmendia J, et al. Modeling Klebsiella pneumoniae Pathogenesis by Infection of the Wax Moth Galleria mellonella. Infect Immun. 2013;81(10):3552-65.

425. Pereira MF, Rossi CC, Seide LE, Martins Filho S, Dolinski C de M, Bazzolli DMS. Antimicrobial resistance, biofilm formation and virulence reveal Actinobacillus pleuropneumoniae strains' pathogenicity complexity. Res Vet Sci. 2018;118:498-501.

426. García-Castellanos R, Marrero A, Mallorquí-Fernández G, Potempa J, Coll M, Gomis-Rüth FX. Three-dimensional structure of MecI. Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. J Biol Chem. 2003;278(41):39897-905.

427. Weiss MS. Global indicators of X-ray data quality. Appl Crystallogr. 2000;34:130-5.

428. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science (80-) [Internet]. 2012;336(1):1030-3. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf

ANNEXES

Annex I: Publications associated to this Doctoral Thesis

Received: 6 June 2022 | Revised: 16 August 2022 | Accepted: 19 August 2022

DOI: 10.1002/pro.4427

FULL-LENGTH PAPER

Repositioning small molecule drugs as allosteric inhibitors of the BFT-3 toxin from enterotoxigenic *Bacteroides fragilis*

Ana Jimenez-Alesanco[1,2] | **Ulrich Eckhard**[3] | **Marta Asencio del Rio**[1,4] | **Sonia Vega**[1] | **Tibisay Guevara**[3] | **Adrian Velazquez-Campoy**[1,2,4,5] | **Francesc Xavier Gomis-Rüth**[3] | **Olga Abian**[1,2,4,5]

[1]Institute for Biocomputation and Physics of Complex Systems (BIFI), Joint Unit GBsC-CSIC-BIFI, Universidad de Zaragoza, Zaragoza, Spain

[2]Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, Zaragoza, Spain

[3]Proteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona (IBMB), Higher Scientific Research Council (CSIC), Barcelona, Catalonia, Spain

[4]Instituto de Investigación Sanitaria de Aragón (IIS Aragon), Zaragoza, Spain

[5]Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas Digestivas (CIBERehd), Madrid, Spain

Correspondence

Adrian Velazquez-Campoy and Olga Abian, Department of Biochemistry and Molecular and Cellular Biology, Institute of Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain.
Email: adrianvc@unizar.es and oabifra@unizar.es

Francesc Xavier Gomis-Rüth, Proteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona (IBMB), Higher Scientific Research Council (CSIC), 08028 Barcelona, Catalonia, Spain.
Email: xgrcri@ibmb.csic.es

Funding information

Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas; Fundació la Marató de TV3, Grant/Award Number: 201815; Gobierno de Aragón, Grant/Award Numbers: E25_20R,

Abstract

Bacteroides fragilis is an abundant commensal component of the healthy human colon. However, under dysbiotic conditions, enterotoxigenic *B. fragilis* (ETBF) may arise and elicit diarrhea, anaerobic bacteremia, inflammatory bowel disease, and colorectal cancer. Most worrisome, ETBF is resistant to many disparate antibiotics. ETBF's only recognized specific virulence factor is a zinc-dependent metallopeptidase (MP) called *B. fragilis* toxin (BFT) or fragilysin, which damages the intestinal mucosa and triggers disease-related signaling mechanisms. Thus, therapeutic targeting of BFT is expected to limit ETBF pathogenicity and improve the prognosis for patients. We focused on one of the naturally occurring BFT isoforms, BFT-3, and managed to repurpose several approved drugs as BFT-3 inhibitors through a combination of biophysical, biochemical, structural, and cellular techniques. In contrast to canonical MP inhibitors, which target the active site of mature enzymes, these effectors bind to a distal allosteric site in the proBFT-3 zymogen structure, which stabilizes a partially unstructured, zinc-free enzyme conformation by shifting a zinc-dependent disorder-to-order equilibrium. This yields proBTF-3 incompetent

---

Abbreviations: ANS, 1-anilino-8-naphthalene sulfonic acid; BFT, *Bacteroides fragilis* toxin or fragilysin; CD, catalytic domain; CD, circular dichroism; CDP, conditionally disordered protein; DMEM, Dulbecco's modified Eagle medium; DMSO, dimethyl sulfoxide; DSF, differential scanning fluorimetry; EDTA, Ethylenediaminetetraacetic acid; ETBF, enterotoxigenic *Bacteroides fragilis*; HCV, hepatitis C virus; ITC, isothermal titration calorimetry; MP, metalloprotease; NEAA, non-essential amino acid; NS3, nonstructural protein 3; NTBF, non-enterotoxigenic *Bacteroides fragilis*; PBS, phosphate-buffered saline; PD, prodomain; RIPA, radioimmunoprecipitation assay buffer; TSA, thermal shift assay; UV, ultraviolet.

Ana Jimenez-Alesanco and Ulrich Eckhard contributed equally.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

E45_20R; Instituto de Salud Carlos III, Grant/Award Number: CPII13/00017; Ministerio de Ciencia e Innovación, Grant/Award Number: PID2019-107725RG-I00; Secretaría de Estado de Investigación, Desarrollo e Innovación, Grant/Award Number: BFU2016-78232-P; Catalan Agency, Grant/Award Number: 2017SGR3; Diputación General de Aragón Predoctoral Research Contract 2019; European Union, Grant/Award Numbers: PI21/00394, PI18/00349

Review Editor: Aitziber Cortajarena

for autoactivation, thus ablating hydrolytic activity of the mature toxin. Additionally, a similar destabilizing effect is observed for the activated protease according to biophysical and biochemical data. Our strategy paves a novel way for the development of highly specific inhibitors of ETBF-mediated enteropathogenic conditions.

KEYWORDS

allosteric inhibitors, *Bacteroides fragilis*, drug repurposing, fragilysin, molecular experimental screening, zinc-dependent metalloproteases

1 | INTRODUCTION

The human colon hosts the largest population of bacteria of the human body. Notably, approximately 25% of its species belong to the genus *Bacteroides*, which are anaerobic, bile-resistant, and non-spore-forming Gram-negative rods.[1] Although *Bacteroides* are normally commensals, several species can cause severe infections in the intra-abdominal space, for example after rupture of the gastrointestinal tract or intestinal surgery.[2,3] After infiltration of the normally sterile peritoneal cavity by gut bacteria, first aerobes such as *Escherichia coli* dominate the infection. However, once sufficient oxygen has been depleted, *Bacteroides* typically dictate the chronic stage. Crucially, untreated *Bacteroides fragilis* infections lead to mortality rates of ~60%.[4,5]

Among *Bacteroides*, *B. fragilis* accounts for only 0.5% of the human colon microbiome, but it is the most frequent anaerobic isolate in clinical specimens,[1] among which enterotoxigenic *B. fragilis* (ETBF) strains are considered the most virulent.[2,6] Remarkably, the only recognized virulence factor specific to ETBF when compared to its nontoxigenic counterparts (NTBF) is a unique 20-kDa zinc-dependent metallopeptidase (MP) dubbed fragilysin or *B. fragilis* enterotoxin (BFT),[7,8] which is not found in any other organism, and is only very distantly related to ADAM/adamalysin MPs.[9] The enzyme is encoded as three isotypes of >90% sequence identity (BFT-1, -2, and -3)[7,8,10] by a 6-kb pathogenicity island (*Bf*PAI) unique to ETBF,[8,11] which is flanked by conjugative transposons. Thus, it may be transmitted to NTBF strains by lateral gene transfer.[12]

BFT can degrade the *zonula adherens* tight junctions of the intestinal epithelium by cleaving E-cadherin,[7,13,14] thereby causing (a) delocalization of other tight-junction proteins, (b) loss of cell adhesion, (c) rearrangement of actin cytoskeleton, (d) nuclear translocation of β-catenin, (e) secretion of inflammatory signaling molecules, and (f) loss of fluids, which collectively cause diarrhea and other related pathologies.[3] As a consequence, patients with ETBF exhibit an increased risk for inflammatory bowel disease and colorectal cancer.[3,15] Indeed, BFT expression is associated with early-stage carcinogenic lesions,[16] and it has been shown that BFT activates both the NF-κB and β-catenin/Tcf signaling pathways, thereby upregulating the c-Myc oncogene and the proinflammatory cytokine IL-8, which are both associated with colorectal cancer.[17–20] To tackle *Bacteroides* infections, β-lactams co-administered with β-lactamase inhibitors, carbapenems, clindamycin, and metronidazole are frequently prescribed, the latter two often in combination with fluoroquinolones.[21] However, several *B. fragilis* strains are intrinsically resistant to several classes of structurally unrelated antibiotics.[22–25] Thus, specific inhibition of BFT represents an attractive novel route to combat ETBF-mediated pathogenicity in inflammatory bowel disease and colorectal cancer without disturbing the commensal microbiota.

Here, we describe a comprehensive drug discovery approach to target (pro)BFT-3. Starting from a collection of 1,120 drugs approved by the United States Food and Drug Administration, we identified compounds capable of specifically binding to the catalytic domain (CD) of (pro) BFT-3 by a molecular experimental screening based on differential scanning fluorimetry (DSF), which assessed ligand-induced stabilization of the protein against thermal denaturation through fluorescence emission spectroscopy. For this, we applied a strategy specially designed for zinc-dependent proteins, which assumes the existence of a physiologically relevant conformational state in the absence of the zinc cofactor. This strategy has proven successful previously for the identification of allosteric inhibitors of the hepatitis C virus (HCV) NS3 protease.[26] Hit compounds were subsequently assessed for target engagement, cytotoxicity, and biological efficacy using both in vitro assays and cell-based E-cadherin processing assays. Furthermore, we determined five crystal structures of

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

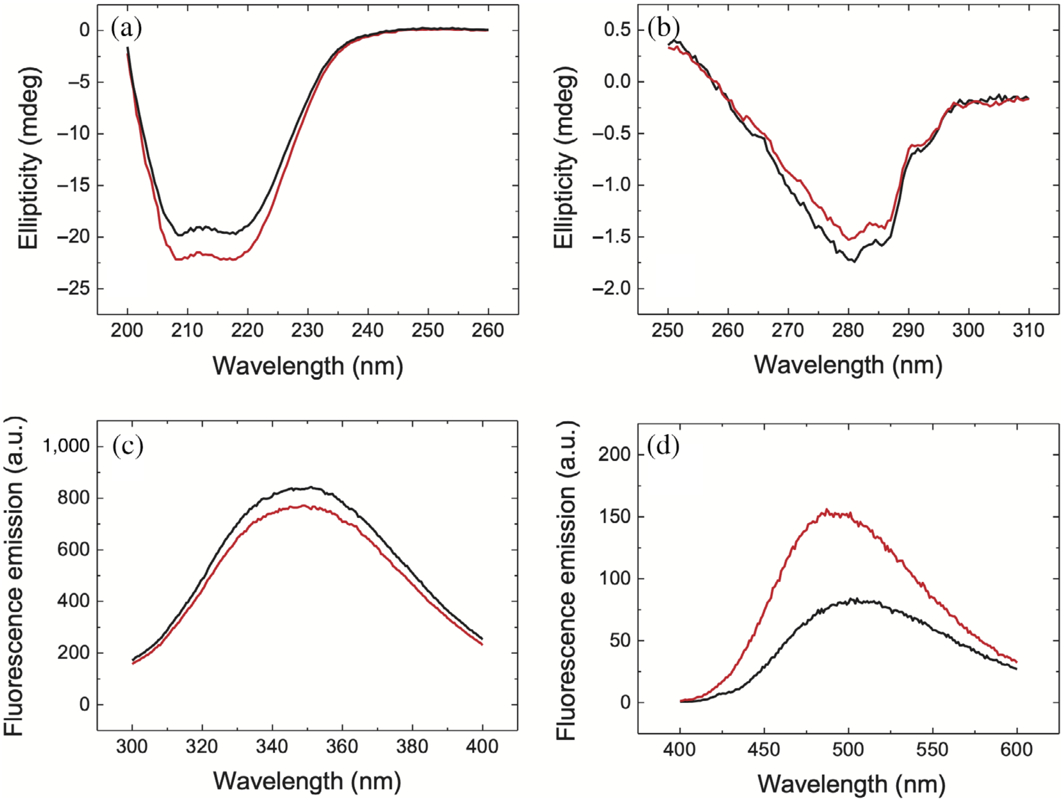


FIGURE 1 Spectroscopic characterization of proBFT-3. Spectra were recorded in the presence (black) and absence (red) of zinc. (a) Far-UV CD spectra and (b) near-UV CD spectra. Raw ellipticity data obtained with 10 μM protein are shown. (c) Fluorescence spectra (excitation wavelength, 280 nm). Data obtained with 2 μM protein. (d) 1-anilino-8-naphthalene sulfonic acid (ANS) fluorescence spectra (excitation wavelength, 370 nm) with proBFT-3. Data obtained with 2 μM protein and 100 μM ANS are shown

proBFT-3, unbound and complex with three of these inhibitors, which uncovered a hitherto unknown distal exosite on the surface opposed to the active-site cleft that crucially impacts the active-site geometry and, thus, activation of proBFT-3 through a unique mechanism of allosteric inhibition. Additionally, as the exosite is located within the CD, a similar destabilizing and inhibitory effect was observed for activated BFT-3. Therefore, because the inhibitors do not bind into the active site, they allosterically affect both proBFT-3 and BFT-3. Importantly, the identified compounds can be either considered straightforward for drug repurposing in combination therapies, or for lead optimization to obtain even more potent compounds against ETBF.

2 | RESULTS

2.1 | proBFT-3 is a zinc-dependent conditionally disordered protein

We employed several biophysical techniques, including circular dichroism (CD), DSF, and both intrinsic and extrinsic fluorescence probes to assess the zinc-dependence of proBFT-3. As expected from the published structure, which evinces an α-helical content of ∼25%,[9] the far-UV CD spectrum revealed a marked negative band at around 220 nm (Figure 1a). Unexpectedly, this peak increased by ∼11% upon ethylenediaminetetraacetic acid (EDTA)-mediated zinc removal while no other major changes were observed. Importantly, the near-UV CD spectrum was non-zero (Figure 1b), which is indicative of a folded structure, and showed small changes upon zinc removal. The intrinsic fluorescence signal at 350 nm was reduced by ∼10% (Figure 1c), pinpointing conformational changes in the local microenvironment of at least one of the four tryptophan residues in the CD of proBFT-3. Next, we investigated the impact of zinc dissociation using the extrinsic fluorescent probe 1-anilino-8-naphthalene sulfonic acid (ANS) as a hydrophobic reporter (Figure 1d). In the zinc-bound form, a fluorescence spectrum with a maximum intensity at around 505 nm was recorded, which is lower than the typical value of 535 nm for free ANS in an aqueous solvent and rather conforms to protein-bound ANS.[27] Upon zinc removal, this blue-shift was enhanced to ∼490 nm, and

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

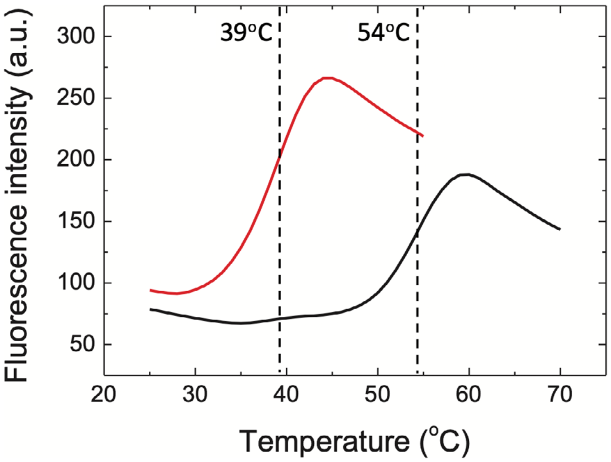


FIGURE 2 Thermal stability of proBFT-3 assessed by differential scanning fluorimetry (DSF). Thermal unfolding of proBFT-3 was monitored by DSF in the presence (black) and absence (red) of zinc. Thermograms were obtained with 2 μM protein and 5X SYPRO Orange. For zinc-free proBFT-3, EDTA was added. For zinc-bound proBFT-3, no exogenous zinc was added; therefore, zinc concentration was that corresponding to the endogenous zinc accompanying proBFT-3 along the purification. Thus, a minor population of zinc-free enzyme could be observed (small positive slope at 39°C in the black line). The zinc bound to proBFT-3 increased the T_m from 39 to 54°C ($\Delta T_m = 15$°C)

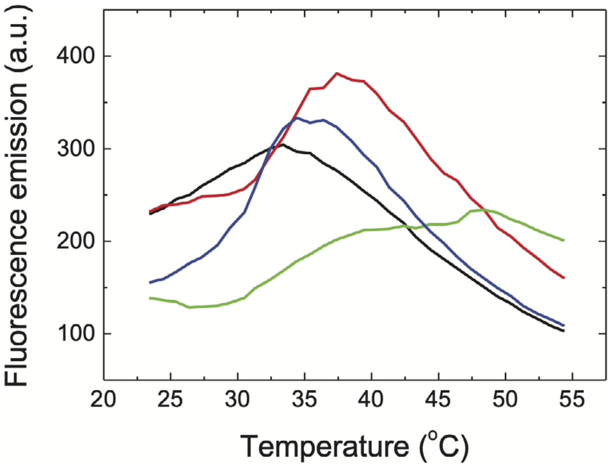


FIGURE 3 Experimental screening of compounds by thermal shift assay. Potential ligands of zinc-free proBFT-3 were identified as those molecules increasing the protein unfolding temperature by at least 3°C. Raw fluorescence intensity data as a function of temperature is shown for compound-free protein and for protein in the presence of selected compounds. Color coding: no compound (black), C-4 (red), C-9 (blue), and C-10 (green)

the recorded fluorescence intensity nearly doubled, indicating a potential conformational change accompanying the cofactor dissociation together with an increase in solvent-exposed protein surface.

In addition, thermal denaturation monitored by DSF showed that zinc-bound proBFT-3 unfolds in a single highly-cooperative transition, which indicates a high folding barrier and a sharp unfolding equilibrium (Figure 2). The unfolding temperature T_m of 54°C was in good agreement with the previously reported value of 56°C[9] and an unfolding enthalpy $\Delta H(T_m)$ of 101 kcal/mol was estimated. Remarkably, zinc depletion caused a substantial reduction of the protein stability as revealed by the unfolding temperature and the associated unfolding enthalpy (39°C and 74 kcal/mol).

2.2 | Identification of ligands targeting the zinc-free partially disordered proBFT-3 state

In a previous study, we identified several small molecules as allosteric inhibitors of HCV NS3 protease, which stabilized the zinc-free—and thus an inactive and partially unfolded protein conformation—through a novel mechanism.[26] Leveraging the overall similarity of BFT-3 and NS3 as zinc-dependent MPs, we implemented a similar experimental strategy for targeting proBFT-3 by screening an FDA-approved chemical library to identify compounds capable to (a) induce stabilization of proBFT-3 against thermal unfolding, and (b) trap proBFT-3 in its zinc-free and thus inactive conformation.

Thermal unfolding in the presence and absence of small molecule compounds was monitored by following the emission signal of the extrinsic fluorescent probe SYPRO Orange, and using the Prestwick Chemical Library, a collection of 1,120 FDA-approved drugs with known therapeutic indication, good bioavailability, and safety in humans, and covering a large chemical and pharmacological space, as the screening set. Initial hits were identified as those compounds substantially increasing the unfolding temperature of zinc-free proBFT-3 at least by 3°C (Figure 3), and selected for further testing (Table 1). The known therapeutic indication for each of the compounds is reported in Table S3.

2.3 | Target engagement of selected compounds

As ligand-induced stabilization of proBFT-3 represents indirect evidence for interaction and there is no direct correlation between protein stabilization extent (i.e., ΔT_m) and ligand affinity (K_a), we determined the dissociation constants for selected compounds by isothermal titration calorimetry (ITC) to assess target engagement. Calorimetric assays were performed for all

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Dissociation constants of the proBFT-3-compound complexes obtained by ITC at 25°C and pH 7.4

Compound	Zn (±)	K_a (M^{-1})	K_d (μM)	ΔH (kcal/Mol)	n
C-4	−Zn	1.3×10^5 (1.2×10^5–1.4×10^5)	7.8 (7.3–8.2)	−5.7 (−5.9 to −5.5)	0.76 (0.72–0.74)
	+Zn	5.0×10^4 (4.4×10^4–5.6×10^4)	20 (18–23)	−9.1 (−10.5 to −8.1)	0.73 (0.71–0.80)
C-9	−Zn	1.2×10^5 (1.1×10^5–1.2×10^5)	8.5 (8.0–8.9)	−6.8 (−7.1 to −6.6)	0.67 (0.66–0.68)
	+Zn	6.9×10^4 (6.3×10^4–7.6×10^4)	14 (13–16)	−10.2 (−11.2 to −9.4)	0.66 (0.62–0.69)
C-10	−Zn	1.2×10^5 (1.1×10^5–1.2×10^5)	8.7 (8.2–9.2)	−3.6 (−3.7 to −3.5)	0.83 (0.82–0.85)
	+Zn	4.6×10^4 (3.7×10^4–5.6×10^4)	22 (18–27)	−7.5 (−8.8 to −6.1)	0.67 (0.55–0.75)

Note: The uncertainty in the estimation of the binding parameters is reported as the confidence interval at a statistical significance of 68%, shown in parenthesis below each parameter.

Abbreviations: ITC, isothermal titration calorimetry; K_a, association constant; K_d, dissociation constant; n, binding stoichiometry (or percentage of binding-competent protein); ΔH, binding enthalpy.

compounds with proBFT-3 in the presence/absence of the active-site zinc (Figure 4). Compounds C-4, C-9, and C-10 exhibited dissociation constants in the low micromolar range, and importantly, the presence of zinc caused an approximately two to threefold reduction in binding affinity, which was most pronounced for compounds C-4 and C-10. Interestingly, the binding enthalpies were more favorable for the zinc-bound proBFT-3, suggesting that the entropic contribution of the binding is either less favorable or more unfavorable in the zinc-free protein. Additional experiments done with trypsin-activated BFT-3 in the presence/absence of the active-site zinc provided similar binding affinities and similar influence of the zinc cofactor (see Table S4). Noticeably, the binding enthalpies were more favorable (i.e., more negative) than those for proBFT-3, and, therefore, the binding entropies were less favorable or more unfavorable than those determined for proBFT-3.

2.4 | Selected compounds inhibit BFT-3 in vitro

To evaluate the potential inhibitory effect of the selected compounds on the BFT-3 proteolytic activity, the commercial EnzChek Protease Assay kit was employed. Presence of proteolytic activity leads to the processing of the quenched BODIPY FL casein substrate, causing an increase in fluorescence intensity that is directly proportional to the enzymatic activity (Figure 5). Activity was determined for trypsin-activated BFT-3, for proBFT-3, and for trypsin alone as a positive control to assess its potential contribution to the overall signal. As expected, an increase in fluorescence over time could be observed for BFT-3, but not proBFT-3 (see Figure S1), and in presence of the compounds, a dose-dependent inhibition was detected (Figure 5).

2.5 | Cytotoxicity of the selected compounds

Cytotoxicity was tested on HT-29 and HeLa cells at the maximum compound concentration used in any of the assays, namely 400 μM (Figure 6). Importantly, compounds showed little or no cytotoxic effect, indicating that the cytotoxic concentration, CC_{50}, is significantly larger and that there is a potential therapeutic window for inhibition.

2.6 | Selected compounds inhibit E-cadherin processing

Western blot assays were performed to assess whether the selected compounds were able to hamper E-cadherin hydrolysis in HT-29 cells. Intact E-cadherin was observed as a single band at ∼120 kDa, whereas processed E-cadherin could be observed as a band at ∼80 kDa, which corresponds to the E-cadherin ectodomain.[14] It was established from cytotoxic assays that compounds showed little toxicity and concentrations up to 200 μM are safe in cell assays. Compounds C-4, C-9, and C-10 reduced the levels of the E-cadherin processing (Figure 7), maintaining most of E-cadherin in its unprocessed form (MW 120 kDa). GAPDH was used as a housekeeping gene for comparison purposes.

2.7 | Structure determination of inhibitory complexes

proBFT-3 is a two-lobed ∼42 kDa zinc-dependent MP zymogen that is synthesized as a 397-residue preproprotein, including a signal peptide for secretion, a ∼190-residue pro-domain (PD; A^{18}-R^{211}) for folding and enzyme inactivation, and an approximately equally long CD (A^{212}-D^{397}),

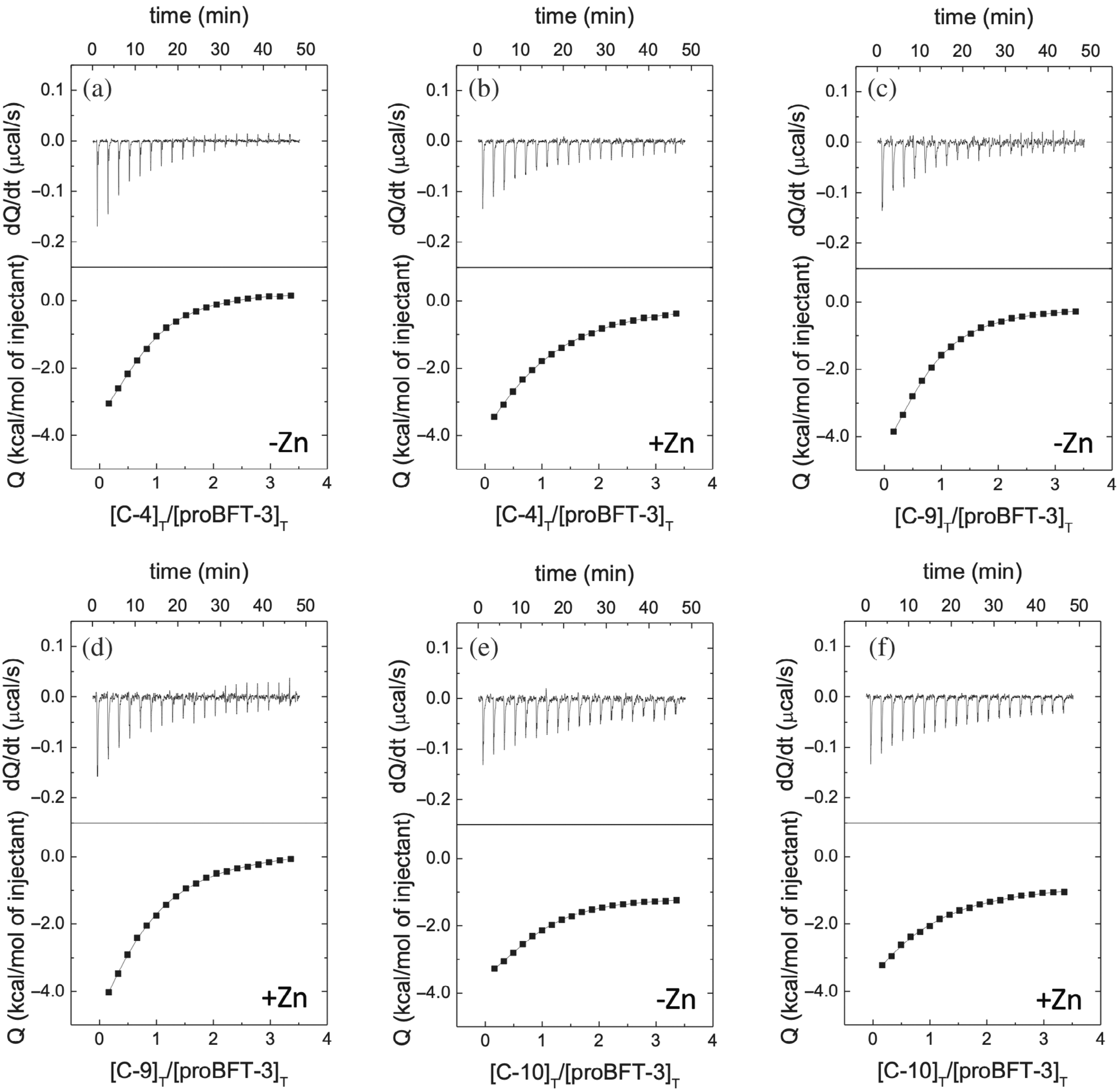


FIGURE 4 Interaction of selected compounds with proBFT-3. Calorimetric titrations for C-4 (a and b), C-9 (c and d) and C-10 (e and f) interacting with proBFT-3, in the absence (left) and the presence (right) of zinc. Thermograms (upper panels; thermal power required to maintain an almost zero temperature difference between sample and reference cell) and binding isotherms (lower panels; ligand-normalized heat effect per peak as a function of the molar ratio) are shown. Nonlinear fits according to a model considering a single ligand binding site (continuous lines) are shown

which belongs to the metzincin clan of metallopeptidases.[9,28] Importantly, the two globular entities of the PD and CD are connected by an inhibitory C-terminal linker segment of the PD (L^{186}-R^{211}), which runs along the active-site cleft and binds to the CD in the opposite direction of a substrate. Via a so-called "aspartate-switch mechanism,"[9] D^{194} of the linker binds to the catalytic zinc ion, thereby displacing the active-site water and rendering the protease inactive in its zymogenic form. The catalytic site is built by a consensus zinc-binding motif, namely HE^{349}XXHXXGXXH, containing three zinc-binding histidine residues and a glutamate residue as the general base/acid for catalysis. The active site is further complemented by a loop segment and a tight 1,4-β-turn centered on M^{366}, the so-called "Met-turn," which forms a hydrophobic basement just underneath the catalytic zinc ion (Figure 8).

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

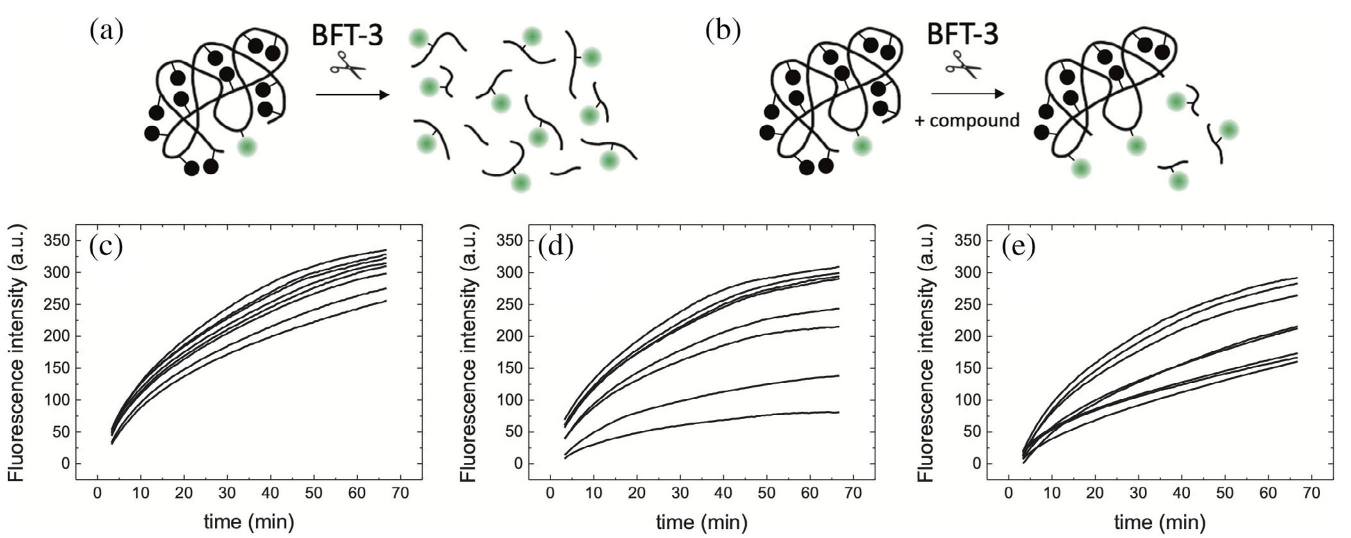


FIGURE 5 Proteolytic activity of BFT-3 measured in vitro, and inhibitory effect of selected compounds. (a and b) BODIPY FL casein was employed as a fluorescent substrate for trypsin-activated BFT-3. Loss of reciprocal fluorophore quenching due to proteolytic processing by BFT-3 results in an increase of fluorescence over time. (a) Maximal activity is observed with BFT-3; (b) Addition of an inhibitor compound reduces the hydrolytic rate, observing a lower fluorescence signal. (c–e) Increasing the concentration of the compounds (twofold serial dilutions from 1 to 0 mM) decreased the activity of BFT-3: (c) C-4, (d) C-9, and (e) C-10. While proBFT-3 showed no activity, trypsin yielded some signal at the used concentrations, but much lower than that of BFT-3, and its contribution was subtracted from the total signal of the trypsin-activated BFT-3 (Figure S1)

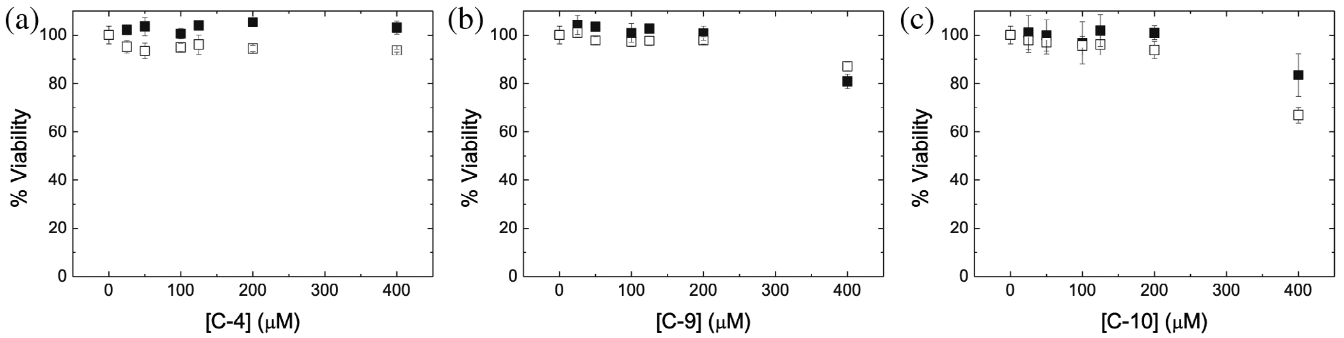


FIGURE 6 Cytotoxicity of compounds against HT-29 and HeLa cell lines. Cell viability values for HT-29 (open squares) and HeLa cells (filled squares) were determined by CellTiter 96® assay after 48 hr of incubation with increasing concentrations of (a) C-4, (b) C-9, and (c) C-10. All data are presented as the average ± standard deviation of three biological replicates, performing the assay twice with technical replicates

We tested the most promising candidates from our biophysical and biochemical characterization for co-crystallization, namely C-4, C-9, and C-10, and succeeded for all three of them. Notably, all compounds bound in similar orientation into a prominent pocket at the top rear of the CD, where the inhibitors inserted wedge-like between the end of the so-called “adamalysin helix” (α5) and the first three strands (β12, β13, β14) of the β-sheet of the upper subdomain (cf. left and a middle panel of Figure 9 and Figure S2). This binding was found consistently in the two protomer complexes present in all crystallographic asymmetric units. Importantly, while for C-9 (307.35 Da) and C-10 (302.29 Da) one inhibitor moiety was found at the exosite with interaction areas of 382 and 364 Å^2, respectively, two molecules of the smaller C-4 (261.26 Da) were bound, thereby increasing the interface from 304 Å^2 for a single C-4 molecule to 497 Å^2. However, while the first C-4 molecule is deeply buried in the protein moiety, with only 14% of its surface being solvent accessible, the second one is found further outside with 43% of its surface solvent accessible. In comparison, for C-9 and C-10, 76% of the surfaces were buried by proBFT-3, which is highly similar to the combined 71% for the two C-4 molecules. However, our ITC data

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

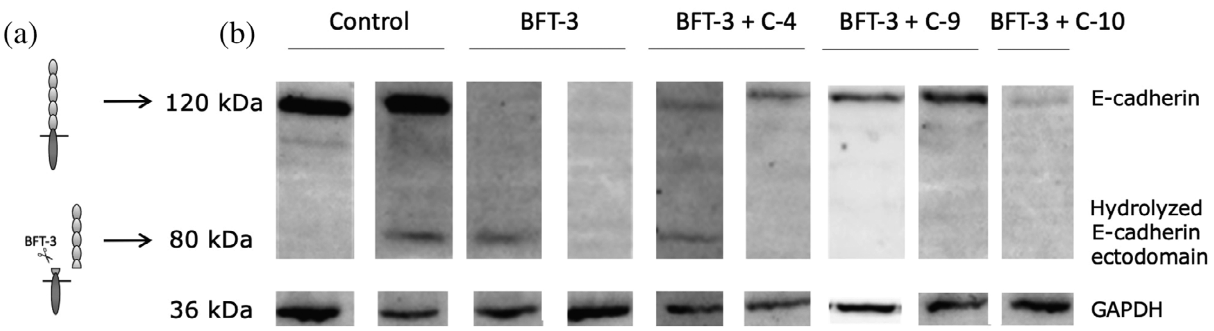


FIGURE 7 Proteolytic activity of BFT-3 and inhibitory effect of selected compounds monitored in a cell-based assay. (a) BFT-3 processing of E-cadherin (MW 120 kDa) results in the release of its ectodomain (MW 80 kDa). (b) In the absence of BFT-3, no degradation of E-cadherin was observed, whereas the presence of BFT-3 led to complete degradation of full-length E-cadherin, while the ectodomain was still partly detectable. The presence of compounds C-4, C-9, and C-10 at 120 μM inhibited E-cadherin processing, as evident by detected western blot bands at 120 kDa. While proBFT-3 was added to the assay, proteases present in the cell culture medium, as well as self-activation, result in activated BFT-3 protein under these conditions

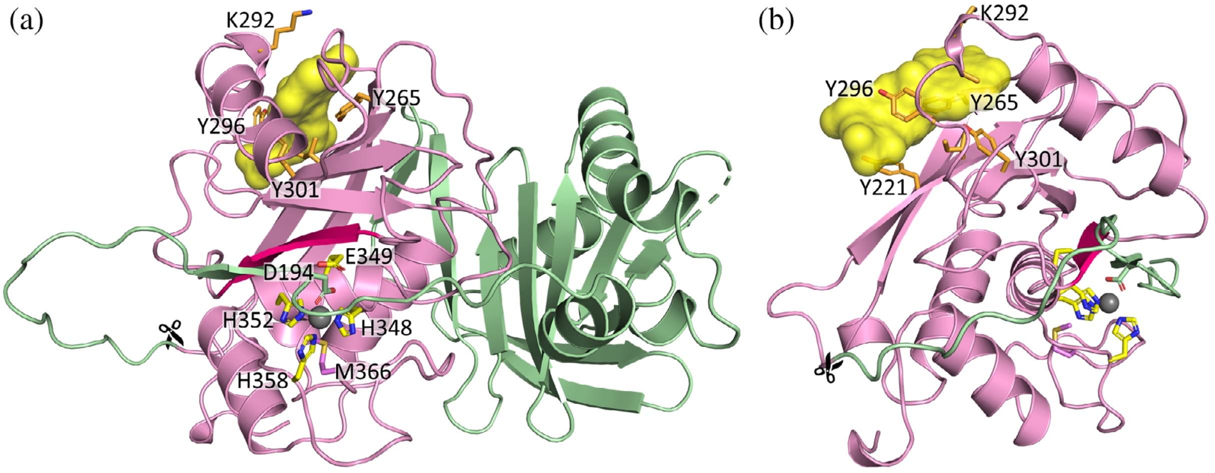


FIGURE 8 Crystallographic structure of proBFT-3 in complex with an identified inhibitor. (a) Overall structure in front view and (b) side view of proBFT-3. The prodomain (PD) and catalytic domain (CD) of unbound proBFT-3 (PDB 7PND) are shown as green and pink ribbons, with α-helices and β-strands as coiled ribbons and flat arrows, respectively. The active-site residues are depicted as sticks with yellow carbons, the catalytic zinc as a gray sphere, and the M^{366} of the hydrophobic basement is shown with violet carbon atoms. Of note, the linker segment of the PD binds to the CD in a reverse direction compared to a substrate, and the aspartate-switch D^{194} is shown in stick mode with green carbons. The exosite responsible for inhibitor binding is located at the back rear of the CD, ∼25 Å away from the active site, and is depicted as a yellow semi-transparent surface resulting from the overlaid small-molecule inhibitors identified in this study extracted from PDB entries 7POL, 7POO, 7POQ, and 7POU, after superposing the respective proBFT-3 moieties. Key inhibitor-binding residues are shown as sticks with orange carbons. The substrate-guiding upper-rim strand of the CD is colored in hot pink, and the activation site of native BFT-3 is indicated with a scissor. The polypeptide chain is interrupted for segment S^{161}-G^{167} of the PD, as indicated by a dashed line. For the side view, PD residues Val^{34}-Asn^{185} were omitted for clarity. Of note, corresponding molecular representations of all five determined X-ray structures, each with two protomers per asymmetric unit, are shown in Figure S2

showed no evidence for the second C-4 molecule, most likely due to the lower inhibitor concentrations used in the calorimetric experiment, implying a much higher affinity for the first binding event.

The main ligand interactions by proBFT-3 are performed with four tyrosines and a lysine residue (K^{292}). Tyrosines Y^{221}, Y^{265}, and Y^{301} are located on β-strands β12, β13, and β14, respectively, while Tyr^{296} is provided by the loop after helix α5. Together, they outline the ligand binding pocket and literally sandwich the bound ligands, with the lysine closing in from the top (Figure 9, left panel). Importantly, while the tyrosines nearly perfectly superimpose in both the non-ligated and complexed structures, the lysine sidechain closes the distance to the opposing side of the exosite cleft by nearly 2 Å (measured between $K^{292}N\zeta$ and $T^{266}C\alpha$; see

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

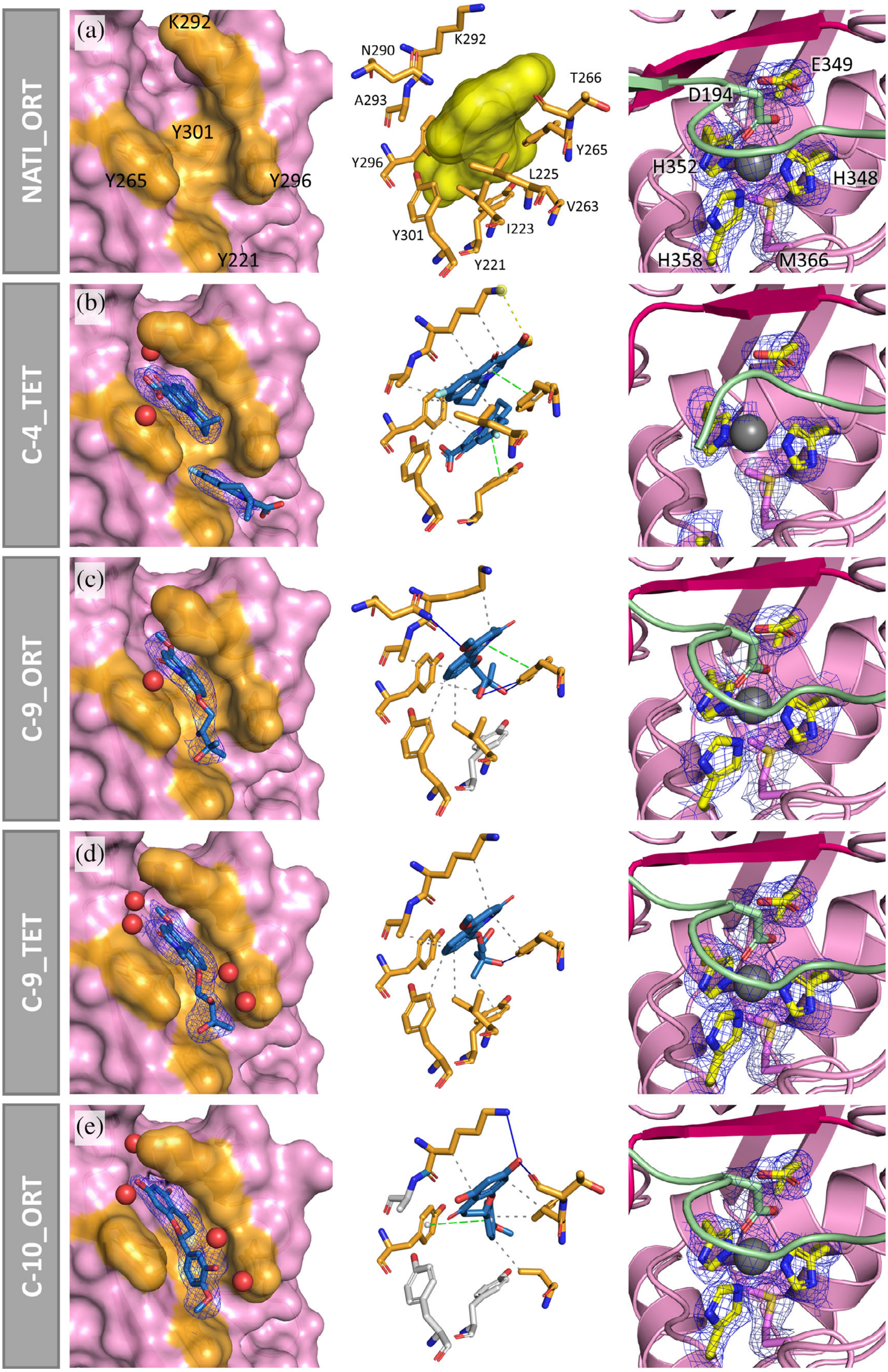

NATI_ORT
C-4_TET
C-9_ORT
C-9_TET
C-10_ORT
(a)
(b)
(c)
(d)
(e)
K292
Y301
Y265
Y296
Y221
N290
K292
A293
T266
Y265
Y296
L225
V263
Y301
I223
Y221
E349
D194
H352
H348
H358
M366

FIGURE 9 Legend on next page.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

e.g., Figure S2). Additionally, while the tyrosines contribute with both hydrophobic and parallel π-stacking interactions to inhibitor binding, K^{292} provides, in addition to hydrophobic interactions, a salt bridge and a hydrogen bond in the C-4 and C-10 complexes, respectively (Figure 9). Overall, the three compounds were found in similar orientation and performing analogous interactions (Figure S4) with the exception of the extended binding interface for C-4 due to the second inhibitory molecule. Ligand-omitted and bulk solvent-excluding mFo–DFc polder maps[29] were calculated using PHENIX,[30] contoured at 3.0 σ in Coot,[31] and are shown for all exosite-binding inhibitory compounds in Figure S5.

Notably, inhibitor binding at the exosite appeared to impact protein crystallization. While the non-ligated proBFT-3 crystalized in the orthorhombic (ORTH) space group $P2_12_12_1$ and diffracted to 1.84 Å resolution, the complex structures with C-4 (1.95 Å), C-9 (1.84 Å), and C-10 (2.03 Å) all crystallized in the tetragonal (TETR) space group $P4_12_12$. For C-9, we further succeeded in obtaining an ORTH structure, but at a significantly lower resolution (2.70 Å). Intriguingly, the building block of both space groups, the crystallographic asymmetric unit, is protein-wise identical in both space groups, with two protomers packing back-to-back and a protein–protein interface area of ∼1,000 $Å^2$. Biocomputational analysis of this surface with Protein Interfaces, Surfaces, and Assemblies software (PISA)[32] indicated this surface was not relevant for dimerization. However, using size-exclusion chromatography, a concentration-independent monomer-dimer equilibrium was observed with significant interface stability based on the immediate reinjection of the dimer peak (Figure S6). For crystallization experiments, both the monomeric and dimeric peaks were concentrated together. Of note, as the interface is largely built by the prodomains, the activated BFT-3 protein is unlikely to dimerize.

Importantly, inhibitor binding to the exosite profoundly impacted the binding of the catalytic zinc despite occurring ∼25 Å apart. While the thermal displacement parameters of the active-site residues were only slightly lower in the inhibitor complexes than in the overall structure (−12%), this effect was more than double in the non-ligated structure (−26%), which indicates that inhibitor binding caused active-site destabilization. Similarly, when looking at the active-site zinc ion, we observed an increase in its thermal displacement parameters upon inhibitor binding, which was typically accompanied by a diminished metal occupancy. This impact was most pronounced in the C-4 complex structure, where the active site, especially H^{358}, is distorted in both chains, and the zinc occupancy is reduced to 0.34 for chain A, and the metal is entirely missing in chain B (Figure 9 and Figure S2, right panels). A similar impact was observed for C-9 in the orthogonal space group, where the occupancy fell to 0.22 and 0.59 in chains A and B, respectively. Compared to non-ligated proBFT-3, a 3.3-fold lower occupancy of the zinc ion (calculated as the zinc b-factor normalized to full-occupancy divided by the overall b-factor of the structure) was identified. Even for C-10, where full zinc occupancy was observed, and for the TETR C-9 complex structure, zinc stability was reduced by ∼20%, despite little effect at first sight. At this point, it is important to note that all crystallization experiments

FIGURE 9 Binding of inhibitors to the newly identified exosite on proBFT-3. Close-up views of the inhibitor-binding exosite (left), the protein-ligand interaction network (middle), and the active site of the catalytic domain (CD) (right) in the distinct structures. Of note, chain A of the crystallographic asymmetric unit is depicted for each of the determined X-ray structures. The corresponding second protein molecule (chain B) is shown in Figure S2 with the same settings and orientations as shown here. Left: Back view onto the ligand-binding exosite. proBFT-3 is shown as a pink semi-transparent surface with key ligand-binding residues shown as sticks with orange carbons. The respective inhibitor is shown in sky-blue and stick mode, with the corresponding electron density map (2Fo-Fc) contoured at 1.0 sigma, and using a carve radius of 2.0 Å. Water molecules in hydrogen bond-forming distance (≤3.4 Å) are shown as red spheres. Middle: Protein-ligand interaction networks at the inhibitory exosite. The orientation is similar to the overview structure in Figure 8, and rotated ∼180° away compared to the view of the figures on the left. Inhibitors are shown in atom-color mode with carbons in sky blue. In (a), the exosite of unligated proBFT-3 is shown, but, as in Figure 8, complemented with the cumulative surface of all crystallized inhibitors in yellow. (b) For C-4 (PDB entry: 7POL), two copies of the inhibitor were found at each exosite. For C-9, two crystal forms were identified: (c) orthorhombic (ORTH; PDB entry: 7POO) and (d) tetragonal (TETR; PDB entry: 7POQ). The ligand interaction network for C-10 (PDB entry: 7POU) is shown in (e). All protein-ligand interactions were calculated using the PLIP webserver 43, and are displayed as gray, green, and yellow dashed lines for hydrophobic interactions, parallel π-stacking, and salt bridges, respectively, while hydrogen bonds are depicted as solid blue lines. Right: Active-site views of proBFT-3 in the non-ligated and inhibitor-complexed structures. A similar orientation as in Figure 8 is shown, with the same key active-site residues depicted in stick mode. The molecular representation was overlaid with the corresponding electron density map (2Fo-Fc) contoured at 1.0 sigma, and using a carve radius of 2.0 Å. Importantly, the active site of C-4-complexed proBFT-3 is completely perturbed, highlighted by a low zinc occupancy and missing electron density for the third proteinaceous zinc ligand (H358), despite the presence of 5 μM $ZnCl_2$ in the protein buffer.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

were carried out in the presence of 5 μM $ZnCl_2$. As a result, zinc occupancy in solution may be even lower. This is consistent with the fact that the tested compounds were first identified against zinc-free (EDTA-treated) proBFT-3, their higher affinity toward the zinc-free conformational state, and the observed competitive interplay between compound binding at the identified exosite and zinc occupation at the catalytic site.

3 | DISCUSSION

Depending on their functional role, protein-bound zinc ions can be classified as either structural (i.e., stabilizing the folded protein) or catalytic (i.e., involved in an enzymatic process). While the former ones are typically bound distal from the active site and coordinated by cysteine residues (e.g., HCV NS3 protease), the latter ones are primarily complexed by aspartates, glutamates, and histidines (e.g., H^{348}, H^{352}, and H^{358} in BFT-3) at the active center. However, the latter may also contribute to the overall stability of the protein in addition to facilitate catalysis. Due to differences in the respective coordination spheres, EDTA readily abstracts zinc from BFT-3 at pH 7, but fails to do so in HCV NS3 at pH >6,[33] despite EDTA having a picomolar binding affinity for zinc.[34,35] Importantly, we detected considerable unfolding in proBFT-3 upon zinc removal, albeit to a lesser extent than in cc NS3,[26] and moderate changes were observed in both the secondary structure content (in fact, an unexpected increase in ellipticity) and the intrinsic tryptophan fluorescence. This may be explained by zinc acting as an anchor between the upper and lower halves of the proBFT-3 CD, so that its depletion leads to increased overall protein flexibility while having little impact on the secondary structure content and the tryptophan environment. Indeed, in the absence of zinc, the solvent-exposed surface area increased significantly, as evidenced by an increase in ANS fluorescence quantum yield. In contrast, both the unfolding stability and cooperativity decreased, as reflected by the much lower thermal stability of the protein (lower unfolding temperature and enthalpy), thus confirming that the catalytic zinc ion serves also a critical structural role in proBFT-3. In fact, based on the thermal stability parameters obtained in DSF (Figure 2), an estimated stabilization Gibbs energy of ~3.6 kcal/mol can be attributed to the zinc interaction, which is significantly higher than the intrinsic stabilization energy of the zinc-free state (~2.5–3.0 kcal/mol), and accounts for ~60% of the total proBFT-3 stabilization energy.

Thus, we postulate that BFT-3 is a conditionally disordered protein (CDP), as previously shown for HCV NS3.[36] CDPs function by temporarily adopting a well-folded structure but are otherwise partially or completely disordered. This conformational switch represents a regulatory element that can be triggered by a post-translational modification, interaction with a biological partner, or a change in environmental conditions (e.g., zinc levels). Notably, CDPs have low structural stability and show properties similar to intrinsically disordered proteins as they populate conformational states separated by small conformational energy gaps. In case of (pro)BFT-3, the solvent and the protein alone seem insufficient to drive the folding toward the structured state. However, upon zinc binding, the required interactions and structural context are available. Importantly, the tight spatial, thermodynamic, and kinetic regulation of the intracellular zinc pool at subnanomolar levels,[37–40] the required proBFT-3 plasticity for secretion into the gut lumen (i.e., the protein can adopt a sufficiently relaxed structure to undergo membrane translocation and refold afterwards), and the rather low intrinsic conformational stabilization Gibbs energy of proBFT-3, support the notion that the partially unfolded zinc-free state of proBFT-3 is indeed of physiological importance. This may add an extra layer of proBFT-3 activity regulation on top of the inhibitory function of the PD. While the partially unfolded state due to zinc scarcity would ensure negligible intracellular proBFT-3 activity, its structural plasticity could expedite both protein secretion and activation—likely by a host protease such as trypsin. Nonetheless, enhanced and durable BFT-3 activity would be attainable in the gut lumen upon zinc binding and stabilization of the protein. This conceivable physiological relevance of proBFT-3's zinc-free state prompted us to search for small molecules capable of binding and stabilizing this partially disordered conformation. This would trap the protein in an inactive state by an allosteric inhibition mechanism where ligand binding modulates the conformational equilibrium between the active and inactive state.[41,42] Such a resulting inhibition mechanism can be classified as either competitive or mixed-inhibition depending on whether the substrate and inhibitor are mutually exclusive (i.e., the substrate and inhibitor solely interact with the folded and partially unfolded protein, respectively), or if a ternary complex is possible and retains reduced but evident activity.

We were able to identify 11 known drugs from the Prestwick Chemical Library as potential BFT-3 ligands using a thermal shift assay (TSA)-based screening procedure under experimental conditions favoring zinc dissociation (i.e., in the presence of the chelating agent EDTA)

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and a semi-quantitative YES/NO readout for hit selection. Subsequently, since TSA analysis only provides indirect evidence for ligand interaction, and as the identified stabilization cannot be directly translated into binding affinities (as it additionally depends on [free] ligand concentration, ligand binding enthalpy, and binding heat capacity[43]), we conducted ITC experiments to assess target engagement and to determine actual binding affinities. Furthermore, we evaluated the biological effect of the compounds using both in vitro and cell-based assays.

While several initial hit compounds lacked adequate binding affinity, C-4, C-9, and C-10 demonstrated significant (pro)BFT-3 binding (i.e., a K_d in the micromolar range) and inhibitory efficacy in both cell-based E-cadherin processing assays and in vitro cleavage studies using BODIPY-casein. Notably, the identified compounds bound to both the proBFT-3 zymogen and the active BFT-3 protein, and interacted with both the zinc-free and zinc-bound form, albeit with reduced affinity for the latter (see Figure 4 and Table 1). Intriguingly, the less favorable Gibbs energy of binding to the zinc-bound enzyme was accompanied by a more favorable enthalpic interaction but which was outweighed by a less favorable (or more unfavorable) entropic term. Importantly, this competitive interplay between the active-site metal (and thus fully folded protein) and the identified inhibitors (which favor the zinc-free partially unfolded state) is also reflected in the reduced zinc occupancy in the determined crystal structures of the complexes—remarkably even despite the presence of 5 μM $ZnCl_2$ during crystallization. The crystal structures also explained the similar affinities to the active and zymogen form as the identified exosite is located on the opposite face of the CD and thus distal to the PD. Crucially, as the identified compounds bind to both zinc-free and zinc-bound BFT-3, their overall inhibitory activity may rely on a combined (a) indirect reciprocal (i.e., substrate and inhibitor reciprocally exclude themselves by preferentially interacting with different BFT-3 zinc-ligation macrostates, namely zinc-bound and the zinc-free BFT-3, respectively) and (b) direct reciprocal competitive effect (i.e., substrate and inhibitor exclude themselves reciprocally by interacting with different conformational microstates of the same macrostate, where the inhibitor distorts the catalytic process and hinders substrate binding and/or catalysis). Consequently, the apparent inhibitory effect will depend on the compound's ability to alter the thermodynamics of the enzyme-substrate interaction (i.e., increase in K_m) and dynamics (i.e., decrease in V_{max}), and the residual activity of the ternary enzyme-substrate-inhibitor complex (ideally negligible). In a simple mixed inhibition scheme, the apparent K_m and V_{max} at a certain inhibitor concentration $[I]$ are given by:

$$K_m^{app} = K_m \frac{1+\frac{[I]}{K_i}}{1+\frac{[I]}{\alpha K_i}}, \quad V_{max}^{app} = V_{max} \frac{1+\beta\frac{[I]}{\alpha K_i}}{1+\frac{[I]}{\alpha K_i}} \tag{1}$$

where K_i and αK_i are the inhibition constants for the substrate-free enzyme (equivalent to the dissociation constant of the enzyme-inhibitor complex) and the substrate-bound enzyme (equivalent to the dissociation constant of the enzyme: substrate-inhibitor interaction, with $\alpha < 1$ and $\alpha > 1$ reflecting positive and negative substrate-inhibitor cooperativity, respectively), and β accounting for the residual activity of the ternary complex. At very high inhibitor concentration, this becomes:

$$K_m^{app} = K_m \alpha, \quad V_{max}^{app} = V_{max} \beta. \tag{2}$$

Consequently, a non-zero β factor results in incomplete inhibition even at high inhibitor concentrations. In fact, this is what we observed despite a clear dose–response effect in the in vitro inhibitory experiments (see Figure 5). However, the heterogeneity of the substrate BODIPY-casein complicates interpretation due to the presence of many different cleavage sites at varying quantities, resulting in a complex behavior of both enzyme and inhibitor. Thus, we verified the inhibitory capacity of the selected compounds in a cell-based assay by monitoring E-cadherin processing. The three compounds (C-4, C-9, and C-10) diminished the efficacy of BFT-3 for cleaving E-cadherin (see Figure 7) at concentrations lower than their estimated CC_{50} (see Figure 6).

Furthermore, additional evidence for target engagement and the allosteric mode of inhibition was obtained by X-ray crystallography. Obviously, crystallization of the partially unfolded zinc-free (pro)BFT-3 would be impossible, potentially explaining the need of $ZnCl_2$ for crystallization. But as the identified compounds also bind to the zinc-bound state of (pro) BFT-3, we eventually succeeded in the determination of the complex structures with all three compounds (C-4, C-9, and C-10). The structures revealed that the ligands bound into the same pronounced pocket at the upper rear of the CD. Notably, despite the exosite being located away from the active site, inhibitor binding leads to a destabilization of the active site, as evidenced by reduced zinc occupancies and increased temperature factors.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To conclude, BFT-3, together with the other two isoforms BFT-1 and BFT-2, is the only known virulence factor in ETBF, causing barrier disruption and inflammation of the colon during infection, and it is overexpressed in the mucosa of inflammatory bowel disease and colorectal cancer patients. Using extensive compound screening, biophysical assays, in vitro and cell-based activity assays, and X-ray crystallography, we identified three FDA-approved small molecule drugs, namely flumequine (C-4), foliosidine (C-9), and hesperetin (C-10), that target and inhibit BFT-3 in a dose-dependent manner. They show great promise to be either directly repurposed for preventive or therapeutic treatment of *B. fragilis* chronic infection, diminishing the risk of inflammation and colorectal cancer development, or, given their small molecular mass, to be further optimized to improve their affinity, selectivity, or bioavailability.

4 | MATERIALS AND METHODS

4.1 | Protein expression and purification

The proBFT-3 enterotoxin precursor (residues A^{18}-D^{397}; UniProt access code O86049) was cloned in the pET-28-based pCri-8a expression vector,[43] which attaches a fusion protein with a TEV-cleavable N-terminal His$_6$-tag.[9] Small-scale *E. coli* BL21 (DE3) cell cultures were grown in LB/kanamycin (50 μg/ml) at 37°C overnight and used to inoculate large-scale 6-L cultures of LB/kanamycin (50 μg/ml), which were incubated at 37°C until $OD_{600} \approx 0.6$. Protein expression was induced with 1 mM isopropyl 1-thio-β-*D*-galactopyranoside at 18°C overnight. Cells were harvested by centrifugation at 4°C for 10 min at 20,000*g* in an Avanti J-26 XP Centrifuge (Beckman Coulter) and resuspended in lysis buffer (500 mM NaCl, 20 mM Tris-HCl, pH 7.4). Cells were lysed by sonication (Sonics Vibra-Cell Ultrasonic Liquid Processor) on ice, after adding 20 U/ml benzonase (Merck-Millipore) and 0.5 mg/ml lysozyme (Carbosynth). Cell debris was removed by centrifugation at 4°C for 10 min at 10,000*g*. Supernatants were clarified by subsequent filtration (0.45 μM-pore membrane) and subjected to affinity chromatography purification in an ÄKTA FPLC System (GE Healthcare Life Sciences) using a cobalt HiTrap TALON column (GE Healthcare Life Sciences) and applying an imidazole gradient (10–250 mM). Purity was assessed by SDS-PAGE, and pure protein fractions were pooled and dialyzed to remove imidazole (150 mM NaCl, Tris-HCl 20 mM, pH 7.4). Protein concentration was quantified using the theoretical extinction coefficient (54,780 M^{-1}/cm) at 280 nm. Zinc-free protein was obtained by adding EDTA at a concentration not higher than 1 mM. Whenever required, mature active BFT-3 was obtained from proBFT-3 by activation with trypsin (PAN-Biotech GmbH) added at a 1:100 mass ratio and incubated at room temperature for 3 hr.

4.2 | CD and fluorescence spectrometry

Circular dichroism spectra were recorded in a thermostated Chirascan spectrometer (Applied Photophysics), using a quartz cuvette of 0.1-cm path length (Hellma Analytics) and selecting a bandwidth of 1 nm, a spectral resolution of 0.5 nm, and a response time of 5 s. Temperature was controlled by a Peltier unit and monitored using a temperature probe. Assays were performed in the far-UV range (190–260 nm). Protein concentration was set to 10 μM.

Fluorescence spectra were collected in a Cary Eclipse spectrofluorometer (Agilent) interfaced with a thermostated multicell holder (Peltier). The slit widths were 5 nm for both excitation and emission wavelengths. An extrinsic probe (8-anilino-1-naphthalene sulfonic acid, ANS) was employed to assess the solvent-exposed molecular surface, using an excitation wavelength of 370 nm and recording the emission spectrum from 400 to 600 nm. Assays were carried out at 25°C in a quartz cell of 1-cm path length (Hellma Analytics). proBFT-3 concentration was set to 2 μM and ANS concentration was set to 100 μM. All measurements were performed in 150 mM NaCl, 20 mM Tris-HCl, pH 7.0, without or with 1 mM EDTA to remove the catalytic zinc from the protein.

4.3 | Differential scanning fluorimetry

The thermal stability of proBFT-3 and its modulation by zinc binding was assessed by DSF in a Mx3005p real-time qPCR device (Agilent). The fluorescence emission intensity of the extrinsic fluorophore SYPRO Orange Protein Gel Stain (Thermo Fisher Scientific) was recorded as a function of temperature in 150 mM NaCl, 20 mM Tris-HCl, pH 7.0, at a scanning rate of 1°C/min using an excitation and an emission wavelength filter of 496 and 610 nm, respectively. These were the closest available to the theoretical values of the fluorophore (491 and 586 nm). SYPRO Orange binds to solvent-exposed hydrophobic patches on the protein surface upon protein unfolding, thusly undergoing a fluorescence quantum yield enhancement with increase of the fluorescence intensity.[43–45] In order to deplete the protein of zinc and get the apozymogen or apoenzyme, 200 μM EDTA was added. Data analysis of the experimental data considered

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a single unfolding transition model and provided relevant thermodynamic stability parameters, viz. the unfolding temperature, T_m, and the associated unfolding enthalpy, $\Delta H(T_m)$.

4.4 | Chemical library

The Prestwick Chemical Library (Prestwick Company), consisting of a set of 1,120 FDA-approved drugs selected for providing large chemical and pharmacological diversity, was supplied with compounds dissolved in 100% dimethyl sulfoxide (DMSO) at a concentration of 4 mM. Information on their bioavailability, as well as toxicity and safety in humans, is available from the manufacturer.

4.5 | Experimental ligand screening

Potential ligands for proBFT-3 were identified from the Prestwick Chemical Library following an experimental screening procedure based on the TSA by DSF, which detects ligand-induced protein stabilization against thermal denaturation,[44–47] in a FluoDia T70 fluorescence microplate reader (Photon Technology International). The procedure was similar to that employed previously for identifying small-molecule inhibitors of *Helicobacter pylori* flavodoxin[48] and the intrinsically disordered DNA-binding protein NUPR1,[49] as well as to find pharmacological chaperones for human phenylalanine hydroxylase.[50] Furthermore, as proBFT-3 is a zinc-dependent protein, we employed a modified version of the screening procedure tailored to the identification of potential ligands of the zinc-free conformational state as previously described for the NS3 protease from hepatitis C virus.[26] Briefly, 100 μl volumes containing 1 μM proBFT-3, 100 μM tester compound, and 1X SYPRO Orange in 150 mM NaCl, 5 mM EDTA, 20 mM Tris-HCl, pH 7.4 were dispensed into 96-well microplates (ThermoFast 96 skirted plates, Thermo Scientific). Each reaction volume contained 2.5% DMSO from the tester stock solutions. Thus, control samples of proBFT-3 with 2.5% DMSO were routinely included in each microplate. Solutions were overlaid with 20 μl mineral oil to prevent evaporation and incubated at 25°C for 30 min. Thermal unfolding curves were registered from 25 to 75°C in 1°C steps by monitoring the fluorescence. We used excitation and emission wavelength filters (470 and 570 nm, respectively) that were the closest available to the theoretical values of the fluorophore (491 and 586 nm). Thermal equilibrium was achieved at each temperature by accounting for an equilibration time of 1 min before each measurement, which corresponded to an operational heating rate of ~0.25°C/min. Hits were identified as those increasing the temperature unfolding temperature T_m at least 3°C compared to the internal controls in each microplate.

4.6 | Isothermal titration calorimetry

The interaction of selected compounds with proBFT-3 was assessed with a high-sensitivity isothermal titration calorimeter Auto-iTC200 (MicroCal, Malvern-Panalytical). Experiments were performed at 25°C in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, and 1 mM EDTA was added for experiments with zinc-free protein. The protein solution at ~20 μM was titrated in the calorimetric cell with compound solution at ~300 μM through a series of 19 injections of 2 μl, with a stirring speed of 750 rpm and a reference power of 10 μcal/s. The heat evolved after each ligand injection was obtained from the integration of the calorimetric signal. The heat due to the binding reaction was obtained as the difference between the reaction heat and the corresponding heat of dilution, the latter estimated as a constant value throughout the experiment, and included as an adjustable parameter in the analysis. Control experiments (compound injected into buffer) were performed under the same experimental conditions in order to observe potential unspecific phenomena (e.g., solution composition mismatches or self-association of compounds). The enthalpy change, ΔH, and the association constant, K_a, of the binding reaction were obtained through nonlinear least-squares regression data analysis of the experimental data applying a model considering a single ligand binding site in the protein. Uncertainties for estimated parameters were calculated according to asymmetric profile likelihood confidence intervals.[51] Experiments were performed in replicates and data were analyzed using in-house developed software implemented in Origin 7 (OriginLab).

4.7 | In vitro inhibition of BFT-3 activity

To evaluate the activity of BFT-3 and the inhibitory effect of the tester compounds, proteolytic assays were performed using BODIPY-casein as the substrate (EnzChek Protease Assay Kit; Thermo Fisher Scientific) according to the manufacturer's instructions. Briefly, proBFT-3 at a concentration of 3 μM in 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, was first activated by adding trypsin (PAN-Biotech GmbH) at a 1:100 M ratio in a final volume of 50 μl and incubated for 3 hr at room temperature. Then, twofold serial dilutions of each compound were made

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

from an initial concentration of 3 mM compound in 7% DMSO, 0.1 mM NaN_3, 10 mM Tris-HCl, pH 7.8, and 50 μl of each dilution were added to the protein sample and subsequently incubated for 1–2 hr. Subsequently, 50 μl substrate were added to each protein-compound sample, and the fluorescence intensity was recorded (excitation/emission: 492/516 nm, the available wavelengths closer to the recommended 505/513 nm for the substrate) for 1 h at 37°C in a Stratagene Mx3005P Real-time PCR instrument (Agilent Technologies). Control experiments featuring proBFT-3 without trypsin activation, trypsin, and buffer alone were performed under the same experimental conditions.

4.8 | Cell viability assays

Cellular cytotoxicity of the selected compounds was assessed in HT-29 (human colon adenocarcinoma) and HeLa (human cervix epithelioid carcinoma) cells obtained from ATCC and maintained in Dulbecco's Modified Eagle's Medium (DMEM; PAN-Biotech GmbH) supplemented with 10% FBS (Fetal Bovine Serum), 1% penicillin/streptomycin, and 1% NEAAs (non-essential amino acids) at 37°C with 5% CO_2. Cells were plated and incubated for 48 hr in 96-well plates (8,000 cells in 100 μl per well for HeLa cells; 9,000 cells in 100 μl per well for HT-29 cells) with FBS-supplemented DMEM without phenol red (PAN-Biotech GmbH). Twofold serial dilutions of compounds were added to the cells up to a maximal concentration of 400 μM and incubated for 48 hr. The cytotoxicity, measured as cell viability, was assayed with the CellTiter 96 AQueous One Solution Cell Proliferation Assay Kit (Promega). To this aim, 20 μl of CellTiter buffer, diluted 1:4 in FBS-supplemented DMEM without phenol red, were added to each well of the 96-well assay plates containing the samples. The plates were incubated at 37°C for 2 hr in humidified atmosphere with 5% CO_2, and absorbance was recorded at 490 nm, with background correction at 800 nm. The readout was directly proportional to the number of living cells in the culture. Each condition was assayed in triplicate and experiments were repeated at least two times.

4.9 | In cell BFT-3 activity inhibition

Western-blot assays were performed to detect E-cadherin in cell samples to monitor the inhibitory capacity of tester compounds on processing by proBFT-3. HT-29 cells were seeded in 24-well plates (90,000 cells in 500 μl per well) with FBS-supplemented DMEM without phenol red. After 48 hr, proBFT-3 was incubated with or without tester compounds in phosphate-buffered saline (PBS) for 3 hr. Then, DMEM without phenol red was added to the sample to a final proBFT-3 concentration of 25 nM. Cell media was removed from each well and the proBFT-3 or proBFT-3/compound solution was added to the cell culture. Of note, in the absence of inhibitor, proBFT-3 is expected to be either activated autoproteolytically or by a secreted protease from the cells. After 24 hr, cell lysates were obtained using cold radioimmunoprecipitation assay buffer (RIPA), which contains inhibitors of proteases and phosphatases. Samples were incubated on ice for 15 min to obtain protein extracts, which were diluted into Laemmli loading buffer. Samples were heated to 95°C for 5 min and subjected to 10% SDS-PAGE in a Mini-PROTEAN system (Bio-Rad) operated at 140 V. Proteins were transferred to a PVDF membrane using Mini-PROTEAN adaptors under wet and cold conditions at 400 mA for 60 min. Membranes were then blocked with 5% non-fat dry milk in Tris-buffered saline (TBS) at room temperature for 1 hr and incubated overnight at 4°C with the respective primary antibody (E-cadherin Antibody or GAPDH Antibody, Table S1), previously diluted 1:2,500 with blocking solution supplemented with 0.1% Tween 20 (TBS-T). After three washes with TBS-T, membranes were incubated with the secondary antibody (previously diluted 1:10,000 with TBS-T), which was marked with Alexa Fluor Plus 647 (Table S1) fluorescence label, at room temperature during 1 hr. After three further washes, the signal was recorded in a ChemiDoc Gel Imaging System (Bio-Rad).

4.10 | Protein crystallization

proBFT-3 protein was concentrated to 10–15 mg/ml using Vivaspin 20 centrifugal concentrators of 10-kDa molecular-mass cutoff (Sartorius, PES), and polished using a Superdex S200 gel filtration column connected to an ÄKTA Purifier 10 (GE Healthcare Life Sciences) with 50 mM NaCl, 20 mM Tris-HCl, pH 8.0, as running buffer. Peak fractions were pooled and re-concentrated to ~10 mg/ml, flash frozen in liquid nitrogen as 50-μl aliquots, and stored at −80°C until further use. High-throughput crystallization screenings were performed using a Phoenix dispenser robot (Art Robbins Instruments) in sitting-drop vapor-diffusion setup by mixing 100 nl of protein with 100 nl of reservoir solution. Commercial crystallization screens including JCSG+, PACT Premier, BCS, and LFS (all from Molecular Dimensions), and PEGRX, SaltRX, and Index Screen (all from Hampton Research) were used in duplicate setups to allow for incubation at 20 and 4°C in Bruker AXS Crystal Farms. Initial crystallization hits were subsequently optimized by hand and upscaled to 0.5-μl drops using standard

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

two-dimensional grid screening in sitting-drop vapor diffusion format.

Suitable crystals of unbound proBFT-3 in a new ORTH space group (NATI_ORTH) and in complexes with inhibitors C-4, C-9, and C-10 in ORTH or TETR space groups were obtained at 4°C from drops of 0.5 μl of protein solution (at 7.9 mg/ml in 50 mM sodium chloride, 20 mM Tris-HCl, pH 8.0, supplemented with 5 μM $ZnCl_2$ and 400 μM of the respective inhibitor) and 0.5 μl of reservoir solution. The final crystallization conditions were 18% PEG 3350, 0.2 M MgHCOO, pH 5.9 for NATI_ORTH; 20% PEG 3350, 0.2 M $(NH_4)_2SO_4$ for the C-4 TETR complex; 14% PEG 3350, 0.2 M $MgCl_2$, and 25% PEG 3350, 0.1 M Bis-Tris, pH 5.5, 0.2 M NH_4CH_3COO, for the C-9 TETR and ORTH complexes, respectively; and 22% PEG 3350, 0.1 M Bis-Tris–HCl, pH 5.5, 0.2 M $(NH_4)_2SO_4$ for the C-10 TETR complex. All crystals typically appeared within 7 days, contained two molecules per asymmetric unit, and were harvested and cryoprotected by soaking for 15–30 s in mother liquor supplemented with 2.5 M L-proline and 1 mM of the respective inhibitor. Of note, all inhibitor stocks were originally at 20 mM in 100% DMSO. Thus, crystallization conditions and cryoprotection buffers contained 2% and 5% DMSO, respectively, in all cases.

4.11 | Structure determination and analysis

Complete diffraction datasets of unbound and inhibitor-bound proBFT-3 crystals were collected at beamlines i04-1 and XALOC of the Diamond and ALBA synchrotrons, respectively. Data were integrated, scaled, merged, and reduced using XDS[52] and XSCALE,[53] and transformed with XDSCONV to the reflection file format MTZ for use by the PHENIX,[30] BUSTER/TNT,[54] and CCP4[55] program packages. Structures were solved by molecular replacement using PHASER[56] with one protomer of Protein Data Bank (PDB) entry 3P24[9] as a search model. Data collection and processing statistics are listed in Table S2. Protein-ligand interactions were analyzed using the PLIP web tool with default settings,[57,58] and both protein–protein and protein-ligand interaction areas were calculated using the PDBePISA webserver.[31] Molecular graphics representations were created using an open-source build of PYMOL version 2.5 Plus: The PyMOL Molecular Graphics System, (Version 2.5 Schrödinger, LLC).[59]

AUTHOR CONTRIBUTIONS

Ana Jimenez-Alesanco: Data curation (equal); formal analysis (equal); investigation (equal); software (equal); validation (equal); writing – review and editing (equal). **Ulrich Eckhard:** Data curation (equal); formal analysis (equal); investigation (equal); software (equal); validation (equal); visualization (equal); writing – review and editing (equal). **Marta Asencio del Rio:** Data curation (equal); formal analysis (equal); investigation (equal); writing – review and editing (equal). **Sonia Vega:** Investigation (equal); writing – review and editing (equal). **Tibisay Guevara:** Investigation (equal); writing – review and editing (equal). **Adrian Velazquez-Campoy:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); methodology (equal); resources (equal); software (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Francesc Xavier Gomis-Rüth:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); resources (equal); software (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Olga Abian:** Conceptualization (equal); formal analysis (equal); funding acquisition (equal); methodology (equal); resources (equal); software (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

This study was supported by Miguel Servet Program from Instituto de Salud Carlos III (grant CPII13/00017), Fondo de Investigaciones Sanitarias from Instituto de Salud Carlos III, and European Union (ERDF/ESF, "Investing in your future") (grant PI18/00349 and PI21/00394), Spanish Ministry of Economy and Competitiveness (grant BFU2016-78232-P), Diputación General de Aragón Predoctoral Research Contract 2019, Diputación General de Aragón Protein Targets and Bioactive Compounds Group E45_20R and Digestive Pathology Group B25_20R, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas—CIBERehd, Spanish Ministry of Science and Innovation (grant PID2019-107725RG-I00), and Catalan Agency (grant 2017SGR3), Fundació "La Marató de TV3" (grant 201815), Secretary of Universities and Research and European Union Horizon 2020 COFUND program, Beatriu de Pinós post-doctoral fellowship. We are grateful to Xandra Kreplin and Joan Pous from the IBMB-CSIC High-Throughput Crystallography Platform for assistance with crystallization experiments, and we would like to thank Diamond Light Source and ALBA Synchrotron for beamtime allocation, and the staffs at beamlines i04-1 and XALOC for superb support with X-ray diffraction data collection.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ORCID

Adrian Velazquez-Campoy https://orcid.org/0000-0001-5702-4538

Francesc Xavier Gomis-Rüth https://orcid.org/0000-0002-6848-6874

REFERENCES

1. Wexler HM. *Bacteroides*: The good, the bad, and the nitty-gritty. Clin Microbiol Rev. 2007;20:593–621.
2. Wu S, Rhee KJ, Albesiano E, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med. 2009;15:1016–1022.
3. Sears CL. Enterotoxigenic *Bacteroides fragilis*: A rogue among symbiotes. Clin Microbiol Rev. 2009;22:349–369.
4. Majid M, Andleeb S. Designing a multi-epitopic vaccine against the enterotoxigenic *Bacteroides fragilis* based on immunoinformatics approach. Sci Rep. 2019;9:1–15.
5. Goldstein EJC. Anaerobic bacteremia. Clin Infect Dis. 1996;23:97–101.
6. Biondi A, Basile F, Vacante M. Familial adenomatous polyposis and changes in the gut microbiota: New insights into colorectal cancer carcinogenesis. World J Gastrointest Oncol. 2021;13:495–508.
7. Franco AA, Mundy LM, Trucksis M, Wu S, Kaper JB, Sears CL. Cloning and characterization of the *Bacteroides fragilis* metalloprotease toxin gene. Infect Immun. 1997;65:1007–1013.
8. Chung G-T, Franco AA, Wu S, et al. Identification of a third metalloprotease toxin gene in extraintestinal isolates of *Bacteroides fragilis*. Infect Immun. 1999;67:4945–4949.
9. Goulas T, Arolas JL, Gomis-Rüth FX. Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. Proc Natl Acad Sci U S A. 2011;108:1856–1861.
10. Kling JJ, Wright RL, Moncrief JS, Wilkins TD. Cloning and characterization of the gene for the metalloprotease enterotoxin. FEMS Microbiol Lett. 1997;146:279–284.
11. Franco AA. The *Bacteroides fragilis* pathogenicity Island is contained in a putative novel conjugative transposon. J Bacteriol. 2004;186:6077–6092.
12. Buckwold SL, Shoemaker NB, Sears CL, Franco AA. Identification and characterization of conjugative transposons CTn86 and CTn9343 in *Bacteroides fragilis* strains. Appl Environ Microbiol. 2007;73:53–63.
13. Wu S, Lim KC, Huang J, Saidi RF, Sears CL. *Bacteroides fragilis* enterotoxin cleaves the *zonula adherens* protein, E-cadherin. Proc Natl Acad Sci USA. 1998;95:14979–14984.
14. Wu S, Rhee KJ, Zhang M, Franco A, Sears CL. *Bacteroides fragilis* toxin stimulates intestinal epithelial cell shedding and γ-secretase-dependent E-cadherin cleavage. J Cell Sci. 2007;120:1944–1952.
15. Wong SH, Yu J. Gut microbiota in colorectal cancer: Mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol. 2019;16:690–704.
16. Purcell RV, Pearson J, Aitchison A, Dixon L, Frizelle FA, Keenan JI. Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia. PLoS One. 2017;12:1–10.
17. Rhee K-J, Wu S, Wu X, et al. Induction of persistent colitis by a human commensal, enterotoxigenic *Bacteroides fragilis*, in wild-type C57BL/6 mice. Infect Immun. 2009;77:1708–1718.
18. Sanfilippo L, Li CK, Seth R, Balwin TJ, Menozzi MG, Mahida YR. *Bacteroides fragilis* enterotoxin induces the expression of IL-8 and transforming growth factor-beta (TGF-β) by human colonic epithelial cells. Clin Exp Immunol. 2000;119:456–463.
19. Kim JM, Cho SJ, Oh YK, Jung HY, Kim YJ, Kim N. Nuclear factor-kappa B activation pathway in intestinal epithelial cells is a major regulator of chemokine gene expression and neutrophil migration induced by *Bacteroides fragilis* enterotoxin. Clin Exp Immunol. 2002;130:59–66.
20. Wu S, Morin PJ, Maouyo D, Sears CL. *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. Gastroenterology. 2003;124:392–400.
21. Cuchural CJ, Tally FP. *Bacteroides fragilis*: Current susceptibilities, mechanisms of drug resistance, and principles of antimicrobial therapy. Drug Intell Clin Pharm. 1986;20:567–573.
22. Shah N, Osmon D, Tande AJ, et al. Clinical and microbiological characteristics of *Bacteroides* prosthetic joint infections. J Bone Jt Infect. 2017;2:122–126.
23. Treviño M, Areses P, Peñalver MD, et al. Susceptibility trends of *Bacteroides fragilis* group and characterisation of carbapenemase-producing strains by automated REP-PCR and MALDI TOF. Anaerobe. 2012;18:37–43.
24. Aldridge KE, Ashcraft D, Cambre K, Pierson CL, Jenkins SG, Rosenblatt JE. Multicenter survey of the changing in vitro antimicrobial susceptibilities of clinical isolates of *Bacteroides fragilis* group, *Prevotella*, *Fusobacterium*, *Porphyromonas*, and *Peptostreptococcus* species. Antimicrob Agents Chemother. 2001;45:1238–1243.
25. Betriu C, Culebras E, Gómez M, López F, Rodríguez-Avial I, Picazo JJ. Resistance trends of the *Bacteroides fragilis* group over a 10-year period, 1997 to 2006, in Madrid, Spain. Antimicrob Agents Chemother. 2008;52:2686–2690.
26. Abian O, Vega S, Sancho J, Velazquez-Campoy A. Allosteric inhibitors of the NS3 protease from the hepatitis C virus. PLoS One. 2013;8:1–10.
27. Hawe A, Sutter M, Jiskoot W. Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res. 2008;25:1487–1499.
28. Arolas JL, Goulas T, Cuppari A, Gomis-Rüth FX. Multiple architectures and mechanisms of latency in metallopeptidase zymogens. Chem Rev. 2018;118:5581–5597.
29. Liebschner D, Afonine PV, Moriarty NW, et al. Polder maps: Improving OMIT maps by excluding bulk solvent. Acta Crystallogr D Struct Biol. 2017;73(Pt 2):148–157.
30. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. Acta Crystallogr Sect D Biol Crystallogr. 2010;66:213–221.
31. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60:2126–2132.
32. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372:774–797.
33. De Francesco R, Urbani A, Nardi MC, Tomei L, Steinkühler C, Tramontano A. A zinc binding site in viral serine proteinases. Biochemistry. 1996;35:13282–13287.
34. Wright DL, Holloway JH, Reilley CN. Heats and entropies of formation of metal chelates of polyamine and polyaminocarboxylate ligands. Anal Chem. 1965;37:884–892.

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

35. Sillen LG, Martell AE, Bjerrum J. Stability constants of metal-ion complexes. London: Chemical Society, 1964.
36. Abian O, Vega S, Neira JL, Velazquez-Campoy A. Conformational stability of hepatitis C virus NS3 protease. Biophys J. 2010;99:3811–3820.
37. Colvin RA, Holmes WR, Fontaine CP, Maret W. Cytosolic zinc buffering and muffling: Their role in intracellular zinc homeostasis. Metallomics. 2010;2:306–317.
38. Outten CE, O'Halloran TV. Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. Science. 2001; 292:2488–2492.
39. Sensi SL, Paoletti P, Bush AI, Sekler I. Zinc in the physiology and pathology of the CNS. Nat Rev Neurosci. 2009;10:780–791.
40. Hara T, Takeda TA, Takagishi T, Fukue K, Kambe T, Fukada T. Physiological roles of zinc transporters: Molecular and genetic importance in zinc homeostasis. J Physiol Sci. 2017;67:283–301.
41. Tsai CJ, Nussinov R. A unified view of 'how allostery works'. PLoS Comput Biol. 2014;10:e1003394.
42. Wyman J, Gill SJ. Binding and linkage: Functional chemistry of biological macromolecules. Mill Valley: University Science Books, 1990.
43. Velazquez-Campoy A, Sancho J, Abian O, Vega S. Biophysical screening for identifying pharmacological chaperones and inhibitors against conformational and infectious diseases. Curr Drug Targets. 2016;17:1492–1505.
44. Goulas T, Cuppari A, Garcia-Castellanos R, et al. The pCri system: A vector collection for recombinant protein expression and purification. PLoS One. 2014;9:e112643.
45. Pantoliano MW, Petrella EC, Kwasnoski JD, et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen. 2001;6:429–440.
46. Matulis D, Kranz JK, Salemme FR, Todd MJ. Thermodynamic stability of carbonic anhydrase: Measurements of binding affinity and stoichiometry using thermofluor. Biochemistry. 2005; 44:5258–5266.
47. Lo M-C, Aulabaugh A, Jin G, et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem. 2004;332:153–159.
48. Cremades N, Velazquez-Campoy A, Martinez-Julvez M, et al. Discovery of specific flavodoxin inhibitors as potential therapeutic agents against *Helicobacter pylori* infection. ACS Chem Biol. 2009;4:928–938.
49. Neira JL, Bintz J, Arruebo M, et al. Identification of a drug targeting an intrinsically disordered protein involved in pancreatic adenocarcinoma. Sci Rep. 2017;7:39732.
50. Pey AL, Ying M, Cremades N, et al. Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. J Clin Invest. 2008;118:2858–2867.
51. Paketurytė V, Petrauskas V, Zubrienė A, et al. Uncertainty in protein–ligand binding constants: Asymmetric confidence intervals versus standard errors. Eur Biophys J. 2021;50:661–670.
52. Kabsch W. XDS. Acta Crystallogr sect D Biol Crystallogr. 2010; 66:125–132.
53. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr Sect D Biol Crystallogr. 2010; 66:133–144.
54. Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr Sect D Biol Crystallogr. 2004;60:2210–2221.
55. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr. 2011;67:235–242.
56. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007;40:658–674.
57. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: Fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015;43:443–447.
58. Adasme MF, Linnemann KL, Bolz SN, et al. PLIP 2021: Expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Res. 2021;49:530–534.
59. DeLano WL. The case for open-source software in drug discovery. Drug Discov Today. 2005;10:213–217.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jimenez-Alesanco A, Eckhard U, Asencio del Rio M, Vega S, Guevara T, Velazquez-Campoy A, et al. Repositioning small molecule drugs as allosteric inhibitors of the BFT-3 toxin from enterotoxigenic *Bacteroides fragilis*. Protein Science. 2022;31(10):e4427. https://doi.org/10.1002/pro.4427

1469896x, 2022, 10, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pro.4427 by Readcube (Labtiva Inc.), Wiley Online Library on [05/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

biomolecules

Article

A Phosphorylation-Induced Switch in the Nuclear Localization Sequence of the Intrinsically Disordered NUPR1 Hampers Binding to Importin

José L. Neira [1,2,*], Bruno Rizzuti [3], Ana Jiménez-Alesanco [2], Martina Palomino-Schätzlein [4], Olga Abián [2,5,6,7,8], Adrián Velázquez-Campoy [2,5,6,7,9] and Juan L. Iovanna [10,*]

1 Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, 03202 Elche, Spain
2 Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Joint Units IQFR-CSIC-BIFI, and GBsC-CSIC-BIFI, Universidad de Zaragoza, 50009 Zaragoza, Spain; ajimenez@bifi.es (A.J.-A.); oabifra@unizar.es (O.A.); adrianvc@unizar.es (A.V.-C.)
3 CNR-NANOTEC, Licryl-UOS Cosenza and CEMIF.Cal, Department of Physics, University of Calabria, Via P. Bucci, Cubo 31 C, 87036 Arcavacata di Rende, Cosenza, Italy; bruno.rizzuti@cnr.it
4 Centro de Investigación Príncipe Felipe, 41930 Valencia, Spain; martina@tinet.org
5 Instituto de Investigación Sanitaria Aragón (IIS Aragón), 50009 Zaragoza, Spain
6 Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd), 28029 Madrid, Spain
7 Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, 50009 Zaragoza, Spain
8 Instituto Aragonés de Ciencias de la Salud (IACS), 50009 Zaragoza, Spain
9 Fundación ARAID, Gobierno de Aragón, 50009 Zaragoza, Spain
10 Centre de Recherche en Cancérologie de Marseille (CRCM), INSERM U1068, CNRS UMR 7258, Aix-Marseille Université and Institut Paoli-Calmettes, Parc Scientifique et Technologique de Luminy, 163 Avenue de Luminy, 13288 Marseille, France
* Correspondence: jlneira@umh.es (J.L.N.); juan.iovanna@inserm.fr (J.L.I.); Tel.: +34-96-6658475 (J.L.N.); +33-(0)4-9182-8803 (J.L.I.)

Received: 23 August 2020; Accepted: 9 September 2020; Published: 11 September 2020

Abstract: Several carrier proteins are involved in protein transport from the cytoplasm to the nucleus in eukaryotic cells. One of those is importin α, of which there are several human isoforms; among them, importin α3 (Impα3) has a high flexibility. The protein NUPR1, a nuclear protein involved in the cell-stress response and cell cycle regulation, is an intrinsically disordered protein (IDP) that has a nuclear localization sequence (NLS) to allow for nuclear translocation. NUPR1 does localize through the whole cell. In this work, we studied the affinity of the isolated wild-type NLS region (residues 54–74) of NUPR1 towards Impα3 and several mutants of the NLS region by using several biophysical techniques and molecular docking approaches. The NLS region of NUPR1 interacted with Impα3, opening the way to model the nuclear translocation of disordered proteins. All the isolated NLS peptides were disordered. They bound to Impα3 with low micromolar affinity (1.7–27 μM). Binding was hampered by removal of either Lys65 or Lys69 residues, indicating that positive charges were important; furthermore, binding decreased when Thr68 was phosphorylated. The peptide phosphorylated at Thr68, as well as four phospho-mimetic peptides (all containing the Thr68Glu mutation), showed the presence of a sequential NN(i,$i+1$) nuclear Overhauser effect (NOE) in the 2D-^{1}H-NMR (two-dimensional–proton NMR) spectra, indicating the presence of turn-like conformations. Thus, the phosphorylation of Thr68 modulates the binding of NUPR1 to Impα3 by a conformational, entropy-driven switch from a random-coil conformation to a turn-like structure.

Keywords: circular dichroism; flexibility; fluorescence; importin; intrinsically disordered protein; isothermal titration calorimetry (ITC); molecular docking; nuclear magnetic resonance (NMR); nuclear protein 1 (NPR1); peptide

1. Introduction

Active nuclear translocation happens through importins (also known as karyopherins), together with other proteins such as the GTPase Ran and nucleoporins [1–3]. The classical nuclear import pathway is started by recognition of a nuclear localization sequence (NLS) in the cargo by importin α [4]. The complex cargo importin α binds to importin β; then, this complex goes through the nuclear pore complex (NPC). The GTPase Ran dissociates the ternary complex within the nucleus by interacting with importin β, and both importins α and β are recycled back to the cytoplasm [4]. The human genome encodes seven isoforms of importin α, with three subtypes [4–6]. These isoforms have a role in cell differentiation, gene regulation [5,7], and even in viral infections, because some viral proteins are recognized by specific importins [8].

Importin α is a modular protein built of α-helix repeat armadillo (ARM) units [1,4]. It has two domains: (i) a N-terminal importin β-binding (IBB) domain, approximately 60-residues-long, which is used for binding to importin β before transport through the NPC, and (ii) a C-terminal NLS-binding motif formed by ten ARM units [9]. Structures of several truncated importin α, without the IBB domain [8,9], have shown that the cargo NLS region binds in a disordered conformation. This interaction occurs at a concave site of the elongated structure, involving ARM motifs 2 to 4 (major site) or 6 to 8 (minor site) for the shortest classical monopartite NLSs or both sets of ARM motifs for the largest bipartite NLS regions. When importin β is not present, the IBB domain, which mimics an NLS region, occupies the ARM motifs involved in NLS recognition [9]. This intramolecular interaction has an autoinhibitory role, and it is thought to be relevant in cargo dissociation in the nucleoplasmic side [9].

Intrinsically disordered proteins (IDPs) do not have a unique stable conformation, resulting in a dynamic conformational ensemble that is reflected in a high structural flexibility. They are involved in cell cycle control, signaling, molecular recognition, replication, and transcription processes [10–13]. The discovery of IDPs has shown that protein biological activity is possible even without a well-defined structure [12–14] but, rather, with an extreme structural flexibility. However, IDPs may have a propensity to adopt structures at the local level; this acquisition of local order can be achieved by, among other factors, post-translational modifications [14]. Such modifications, in turn, can widen their biological functions [11,15]. NUPR1 (UniProtKB O60356) is an 82-residue-long (8 kDa), highly basic, monomeric IDP that is overexpressed during the acute phase of pancreatitis [16,17] and in almost any, if not all, cancer tissues [18]. Its exact functions are unknown, but NUPR1 is a key element in the cell-stress response and cell-cycle regulation [18,19]. Moreover, NUPR1 intervenes in apoptosis through the formation of a complex with the oncoprotein ProTα [20] and in DNA repair [21,22]. In the interactions with all these partners and other synthetic molecules, NUPR1 uses two hotspots around residues Ala33 and Thr68 [22–24]. In addition, NUPR1 has a bipartite NLS region around Thr68, which is fully functional [25]. Thus, even though NUPR1 is a relatively small protein, it might require the assistance of the importin system for nuclear translocation due to its unfolded nature and its large radius of gyration, which would be closer to the limit of free diffusion through the NPC. In addition, NUPR1 might require the presence of importins to avoid undesired interactions with other macromolecules in the cytoplasm, due to its basic nature [26].

In this work, we have studied the interaction of human importin α3 (Impα3), also called KPNA4, and that of its truncated species, without the IBB domain (ΔImpα3), to either NUPR1 or peptides encompassing its NLS (NLS-NUPR1). We have chosen Impα3 as a target for NUPR1 because of its larger flexibility when compared with other importins, as concluded by the structural factors from the X-ray data, which confers in it a greater ability to interact with cargos, having a higher variety of conformations [8]. From an experimental pint of view, Impα3 can be also easily expressed and purified for in vitro structural studies [8]. Interestingly, it has also been shown to be crucial in pain pathways [27]. In addition, by studying both importin species (with and without the IBB), we were interested in finding out whether the absence of the IBB domain affected the binding of NLS-NUPR1. The NLS-NUPR1 peptides had mutations at: (i) the two lysines in the sequence (Lys65 and Lys69), which are important for nuclear translocation, according to in vivo studies [25], and (ii) Thr68, where we have either introduced phospho-threonine or, alternatively, we have designed phospho-mimetic mutations (with a glutamic residue). We have used several spectroscopic and biophysical techniques—namely, steady-state fluorescence, circular dichroism (CD), nuclear magnetic resonance (NMR), isothermal titration calorimetry (ITC), and molecular docking—to address the binding of the peptides to both importins. Our results indicate that the isolated wild-type (wt) NLS-NUPR1, as well as the mutants, were monomeric and disordered in the solution. The wt NLS-NUPR1 peptide bound to both importins, and the affinity was larger for ΔImpα3 (0.95 μM versus 1.7 μM for Impα3), indicating that the IBB region must have an inhibitory effect; this result is in agreement with other binding studies involving intact, well-folded protein cargos [9], but to the best of our knowledge, this is the first time tested with an IDP. The binding of NLS-NUPR1 peptides to both importins was hampered by removal of either Lys65 or Lys69, and it was almost abolished when Thr68 was phosphorylated or when the phospho-mimetics were assayed. Interestingly enough, the phosphorylated peptide at Thr68 and the four phospho-mimetics showed the presence of turn-like conformations, which were not observed in the wt NLS-NUPR1 peptide or in the Lys65Ala or Lys69Ala mutants. We concluded that the phosphorylation of Thr68 modulates the binding of NUPR1 to importin by a conformational switch from a random-coil to a turn-like conformation.

2. Materials and Methods

2.1. Materials

Isopropyl-β-D-1-tiogalactopyranoside and ampicillin were obtained from Apollo Scientific (Stockport, UK). Imidazole, kanamycin, Trizma base, and His-Select HF nickel resin were from Sigma-Aldrich (Madrid, Spain). Protein marker (PAGEmark Tricolor) and Triton X-100 were from VWR (Barcelona, Spain). Amicon centrifugal devices were from Millipore (Barcelona, Spain), and they had a cut-off molecular weight of 30 or 50 kDa. The rest of the materials were of analytical grade. Water was deionized and purified on a Millipore system.

2.2. Protein Expression and Purification

The His-tagged ΔImpα3 (residues 64-521) was obtained from BL21 (DE3) cells as described [8]. The DNA of the codon-optimized, intact Impα3 with a His-tag at the N terminus was synthesized by NZYtech (Lisbon, Portugal) and cloned into the pHTP1 vector (with kanamycin resistance). Expression and purification of Impα3 were carried out as those for ΔImpα3 in the same *Escherichia coli* strain. Concentration of both species was determined from their six tyrosines and six tryptophans [28].

2.3. Design and Synthesis of the Peptides

The peptides were synthesized by NZYtech with a purity of 95%. The peptides comprised the NLS region of NUPR1 (Table 1); peptides were named with the accompanying name within parenthesis for each sequence, as reported in Table 1. All peptides were acetylated and amidated at the N and C termini, respectively, to avoid fraying effects. As the wt NLS had no tyrosine, we introduced one at the N terminus to allow for absorbance measurements [28]. We synthesized eight peptides with different mutations, with the following rationale: (i) we studied the importance of positions Lys65 and Lys69 in the binding to both importins by mutating the two positions to alanine, (ii) we mutated Thr68 to the glutamic T68E peptide to have a phosphomimic at this position, (iii) we combined this mutation at

Thr68 with either of the other two as double mutants, as well as to both in a triple mutant, and (iv) we designed the phosphorylated peptide at position Thr68 (pT68 peptide) to study the effects of this single post-translational modification.

Table 1. Hydrodynamic properties of the nuclear localization sequence (NLS) NUPR1 peptides.

Peptide [a]	***D* (cm^2 s^{-1}) × 10^6 (*R*$_h$, Å)** [b]	***R*$_h$, Å** [c]
YT^{54}NRPSPGGHERKLVTKLQNSE (wt)	1.85 ± 0.04 (11 ± 1)	13 ± 3
YTNRPSPGGHER**A**LVTKLQNSE (K65A)	1.94 ± 0.08 (11 ± 1)	13 ± 3
YTNRPSPGGHERKLVT**A**LQNSE (K69A)	1.79 ± 0.06 (12 ± 2)	13 ± 3
YTNRPSPGGHERKLV**E**KLQNSE (T68E)	2.17 ± 0.06 (10 ± 1)	13 ± 3
YTNRPSPGGHER**A**LV**E**KLQNSE (K65AT68E)	1.76 ± 0.06 (12 ± 1)	13 ± 3
YTNRPSPGGHERKLV**EA**LQNSE (T68EK69A)	1.87 ± 0.08 (11 ± 1)	13 ± 3
YTNRPSPGGHER**A**LV**EA**LQNSE (K65AT68EK69A)	2.4 ± 0.2 (9 ± 2)	13 ± 3
YTNRPSPGGHERKLV**pT**KLQNSE (pT68)	1.89 ± 0.08 (11 ± 1)	13 ± 3

[a] Mutations with respect to the wild-type sequence are indicated in bold. The last peptide has a phospho-threonine at position 68 (indicated with a "pT"). [b] The R_h was determined from the translational diffusion coefficient of dioxane (R_h = 2.12 Å) added to each sample. [c] Calculated from the scale law: $R_h = (0.027 \pm 0.01)\ \mathrm{MW}^{(0.50 \pm 0.01)}$ [29], where MW is the molecular weight of the peptide. *D*: translational diffusion coefficient.

2.4. Fluorescence

2.4.1. Steady-State Fluorescence

Fluorescence spectra were collected on a Cary Varian spectrofluorometer (Agilent, Santa Clara, CA, USA) with a Peltier unit. The samples were prepared the day before and left overnight at 278 K; before experiments, samples were left for 1 h at 298 K. A 1-cm-pathlength quartz cell (Hellma, Kruibeke, Belgium) was used. Concentrations of the peptides were 10 μM and that of importins was 4 μM. Samples containing the isolated peptide, the isolated importin, and the mixture of both, at those concentrations, were prepared for each peptide and each importin. Experiments were acquired at pH 7.0 in 50-mM phosphate buffer.

Protein samples were excited at 280 and 295 nm (although the samples of the isolated peptides did not show any fluorescence at the latter value). The other experimental parameters and the buffers used have been described elsewhere [30]. Appropriate blank corrections were made in all spectra.

2.4.2. Thermal Denaturations

Thermal denaturations were performed at 60 K/h with an average time of 1 s for all samples. Thermal scans were collected at 315, 330, and 350 nm after excitation at 280 or 295 nm from 298 to 358 K. The rest of the experimental set-up was the same as described above. Thermal denaturations for both importins were irreversible, as well as that of the complexes with any peptide. The apparent thermal denaturation midpoint was estimated from a two-state equilibrium equation as described [30].

2.5. CD

Far-ultraviolet (UV) CD spectra were collected on a Jasco J810 spectropolarimeter (Jasco, Tokyo, Japan) with a thermostated cell holder and interfaced with a Peltier unit at 298 K. The instrument was periodically calibrated with (+)-10-camphorsulphonic acid. A path length cell of 0.1 cm was used (Hellma, Kruibeke, Belgium). All spectra were corrected by subtracting the corresponding baseline. The concentration of each polypeptide was the same used in the fluorescence experiments. The buffer was the same used in the fluorescence experiments.

2.5.1. Far-Ultraviolet (UV) Spectra

Isothermal wavelength spectra of each sample were acquired with six scans at a scan speed of 50 nm/min, with a response time of 2 s and a bandwidth of 1 nm. The samples were prepared the day before and left overnight at 278 K to allow for equilibration. Before starting the experiments, the samples were further left for 1 h at 298 K.

2.5.2. Thermal Denaturations

The experiments were performed at 60 K/h and a response time of 8 s. Thermal scans were collected by following the changes in ellipticity at 222 nm from 298 to 343 K. The rest of the experimental set-up was the same as reported in the steady-state experiments. Thermal denaturations were not reversible for any of the samples, as shown by: (i) comparison of the spectra before and after the heating and (ii) changes in the voltage of the instrument detector [31]. The apparent thermal denaturation midpoint of the samples was estimated as described [30].

2.6. ITC

The experimental set-up and data processing of ITC experiments has been described previously [32]. Calorimetric titrations, performed in an Auto-iTC200 calorimeter (MicroCal, Malvern-Panalytical, Malvern, UK) consisted of series of 19 2-µL injections, with 150 s time spacing and a 750-rpm stirring speed. Impα3 or ΔImpα3 (at 10–20 µM) was loaded into the calorimetric cell and NLS-NUPR1 peptides in the syringe (150–300 µM); all solutions were prepared in buffer Tris 50 mM, pH 8. The temperature for all the experiments was 298 K. The experiments were analyzed by applying a model considering a single ligand binding site (1:1 stoichiometry) implemented in Origin 7.0 (OriginLab, Northampton, MA, USA). The binding affinity (association constant) and the binding enthalpy were estimated through a least-squares nonlinear regression data analysis, from which the Gibbs energy and the entropic contribution to the binding were calculated using well-known thermodynamic relationships. Since the binding stoichiometry is constrained by the model, the parameter n provides a fraction of the active or binding competent protein. Experiments for each peptide and importin species were performed, at least, in duplicates.

2.7. NMR

The NMR experiments were acquired at 283 K on a Bruker 500 MHz Advance III spectrometer (Bruker GmbH, Karlsruhe, Germany) equipped with a triple-resonance probe and z-pulse field gradients. Temperature of the probe was calibrated with methanol [33]. All experiments were carried out at pH 7.2, 50-mM deuterated Tris buffer (not corrected for isotope effects). The spectra were calibrated with TSP ((trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid) by considering pH-dependent changes of its chemical shifts [33].

2.7.1. 1D-^{1}H-NMR (One-Dimensional Proton NMR) Spectra

In all cases, 128 scans were acquired with 16 K acquisition points and using concentrations of 1.0–1.2 mM. Homonuclear 1D-^{1}H-NMR spectra were processed with Bruker TopSpin 3.1 (Bruker GmbH, Karlsruhe, Germany) after zero-filling and apodization with an exponential window.

2.7.2. Translational NMR Diffusion Ordered Spectroscopy (DOSY)

Concentrations of peptides in all DOSY experiments were 120 µM, and 128 scans, where the gradient strength was varied, were acquired for each curve. Translational self-diffusion measurements were performed with the pulsed gradient spin-echo sequence in the presence of 100% D_2O. Experimental details have been described elsewhere [30]. Briefly, the gradient strength was varied in sixteen linear steps between 2% and 95% of the total power of the gradient coil. The gradient strength was previously calibrated by using the value of the translational diffusion coefficient, D, for the residual proton water line in a sample containing 100% D_2O in a 5-mm tube [34]. In our experiments for each peptide, the duration of the gradient was 2.25 ms, the time between the two pulse gradients in the pulse sequence was set to 200 ms, and the recovery delay between the bipolar gradients was set to

100 μs. The methyl groups with signals between 1.0 and 0.80 ppm were used for integration. Fitting of the exponential curves obtained from experimental data was carried out with KaleidaGraph (Synergy Software, Version 3.5), as described [30]. A final concentration of 1% of dioxane, which was assumed to have a hydrodynamic radius R_h = 2.12 Å [34], was added to the solutions of each of the peptides to have a comparison for estimating their sizes.

2.7.3. 2D-^{1}H-NMR Spectra

Two-dimensional spectra in each dimension were acquired in phase-sensitive mode by using the time-proportional phase incrementation technique (TPPI) and a spectral width of 7801.69 Hz [35]; the final concentration was the same used in the 1D-^{1}H-NMR experiments. Standard total correlation spectroscopy (TOCSY) (with a mixing time of 80 ms) [36] and nuclear Overhauser effect spectroscopy (NOESY) experiments (with a mixing time of 250 ms) [37] were performed with a data matrix size of 4K × 512. The DIPSI (decoupling in the presence of scalar interactions) spin-lock sequence [38] was used in the TOCSY experiments with 1 s of relaxation time. Typically, 64 scans were acquired per increment in the first dimension, and the residual water signal was removed by using the WATERGATE sequence [39]. NOESY spectra were collected with 96 scans per increment in the first dimension, with the residual water signal removed again by the WATERGATE sequence and 1 s of relaxation time. Data were zero-filled and resolution-enhanced in each dimension, with a square sine-bell window function optimized in each spectrum, baseline-corrected, and processed with Bruker TopSpin 3.1. The ^{1}H resonances were assigned by standard sequential assignment processes [40]. The chemical shift values of H_α protons in random-coil regions were obtained from tabulated data, corrected by neighboring residue effects [41,42] and taking into account the phosphorylation of Thr68 [43,44] for the corresponding peptide.

2.8. Molecular Docking

Molecular docking was performed using AutoDock Vina (Version 1.1.2) [45], largely following a protocol we have previously described for screening NUPR1 sequence fragments [24]. The structure of ΔImpα3 was modeled on the basis of the Protein Data Bank (PDB) entry 5 × 8N [46], which reports the X-ray structure of monomeric Impα1 bound to the NLS of the Epstein-Barr virus EBNA-LP protein. The search volume was centered on the macromolecule and had the size 50 Å × 90 Å × 90 Å, which was sufficient to carry out a blind search on the whole protein surface.

The peptides used in our experiments encompassed residues 53–74 of NUPR1, with a number of rotatable dihedral angles ranging from 85 to 91. Their conformational space was too large to be reasonably treated by molecular docking; therefore, we followed a two-fold approach [47] that consisted in reducing the number of degrees of freedom and using a longer search protocol. The number of rotatable dihedrals was halved by considering the reduced sequence that encompasses residues 63–71 of NUPR1, and therefore, it includes only the core region of the NLS. These shorter peptide sequences were capped with an acetyl and N-methyl group at the two main chain endings, to mimic the fact that they are internal portions of the sequence of the protein, as well as of their full-length parent peptides. An extensive search was performed with very high exhaustiveness, 16 times larger than the recommended default value [48].

3. Results

3.1. The Isolated wt NLS-NUPR1 and Its Mutants Were Monomeric and Disordered in Aqueous Solution

We first determined the conformational propensities of isolated peptides by using CD and NMR. We did not use fluorescence to characterize their conformational features, because the peptides only have a single tyrosine at their N terminus, whose maximum wavelength (~308 nm) does not change under different environments in solutions [49]. The CD spectra of isolated peptides did show an intense minimum at ~200 nm (Figure S1), indicating that they were mainly in a random-coil conformation.

This was further confirmed by 1D-^{1}H-NMR spectra, which showed, for all the peptides, a clustering of the signals of all the amide protons between 8.0 and 8.5 ppm (Figure S2) and grouping of the methyl protons between 0.8 and 1.0 ppm, which is a feature of disordered polypeptide chains [40].

The peptides were monomeric, as concluded from the values of D measured by the DOSYs and the calculation of the estimated R_h from a random-coil polypeptide according to an exponential law [29] (Table 1).

To further confirm the disordered nature of the peptides, we also carried out homonuclear 2D-^{1}H-NMR experiments (Tables S1–S8). For all peptides, NOEs between the H_α protons of Arg56 or Ser58 and the H_δ of the two following residues (Pro57 and Pro59, respectively) were always observed (Figure 1); these findings suggest that the Arg56-Pro57 and Ser58-Pro59 peptide bonds predominantly adopted a trans-conformation in all the peptides (other minor signals were not observed). Two lines of evidence confirmed the disordered nature of the peptides (further pinpointing the findings from far-UV CD (Figure S1) and the 1D-^{1}H-NMR spectra (Figure S2)). First, the sequence-corrected conformational shifts ($\Delta\delta$) of H_α protons [40–44] were within the commonly accepted range for random-coil peptides ($\Delta\delta \leq 0.1$ ppm) (Tables S1–S8). It is interesting to note at this stage that, in the phosphorylated Thr68 of the pT68 peptide, the signals from the H_β protons were downfield shifted when compared to those of the wt peptide (4.58 versus 4.15 ppm, respectively), as well as the chemical shift of the amide proton: 8.62 versus 8.33, respectively (Tables S1 and S3), as it has been reported to occur for phosphorylated threonines [43,44], thus confirming the phosphorylation of this particular threonine and not of the other one in the sequence, Thr54. Second, in any of the peptides, no long- or medium-range NOEs were generally detected but, rather, only strong sequential ones ($\alpha N(i,i+1)$) (Figure 1). Only in the pThr68 peptide and in the four phospho-mimics (T68E, K65AT68E, T68EK69A, and K65AT68EK69A peptides), we observed a weak NOE ($NN(i,i+1)$) between the amide protons of Val67 and Thr68 (Figure S3). This NOE, although weak when compared with the intensity of sequential $\alpha N(i,i+1)$ NOEs, is a fingerprint signature of turn-like conformations [40].

Although there are some isolated short peptides that are partially structured (such as the isolated Ribonuclease S peptide [33,40]), our findings by CD and NMR indicate that the isolated NLS-NUPR1 peptides were mainly disordered in aqueous solution when isolated.

3.2. The NLS-NUPR1 Peptides Bound to Both Impα3 and ΔImpα3

In the present work, we measured the affinity of intact NUPR1 for ΔImpα3, obtaining a value for the dissociation constant of 0.4 μM (Figure S4), and we have previously measured the affinity of intact NUPR1 for Impα3, and a value of 1.4 μM has been obtained (shown in Figure S4; for a comparison, [50]). Furthermore, we tried to dissect the affinity of the NLS region of NUPR1 for Impα3 by using a "divide and conquer" approach with the peptides comprising the region. The interaction between full-length NUPR1 and its mutants with Impα3 and ΔImpα3 was the focus of this study, but instead, we employed NLS peptides to elucidate the binding mechanism to Impα3. The reason behind such an approach relies in the fact that we have observed that, very often, mutations at any place of the polypeptide length of NUPR1 result in a poor expression of the corresponding mutant, and mutations in some positions lead to no expression at all [24].

First, we decided to investigate a possible interaction between the NLS NUPR1 peptides and Impα3 in vitro by using fluorescence and CD. As a representative example, we describe our findings for the wt peptide. We observed changes in the fluorescence spectrum of this peptide after excitation at 280 nm (whereas there were no changes at 295 nm); that is, the additional spectrum obtained from the spectra of isolated wt peptide and either Impα3 or ΔImpα3 was different to those of their respective complexes (Figure 2A). These results indicate that tyrosine residues of at least one of the biomolecules (peptides with either Impα3 or ΔImpα3) were mainly involved in the binding. The changes were small for Impα3, and there were no changes for ΔImpα3; furthermore, thermal denaturations followed by fluorescence did not show a variation in the apparent thermal denaturation midpoint for both Impα3 and ΔImpα3 (Figure S5). On the other hand, the comparison of the additional spectrum and that of the complex obtained by far-UV CD did show differences (both for Impα3 and ΔImpα3), indicating that there were changes in the secondary structure of at least one of the macromolecules upon binding (Figure 2B); however, there were no differences in the determined thermal denaturation midpoint for isolated Impα3 (or ΔImpα3) and that of the complex (Figure S5). It is important to note that the far-UV CD region is sensitive to elements of secondary structures (α-helix and β-sheet); however, local structural elements and nonregular structures might also be present, which could be masked by the presence of long disordered regions. The above results indicate that there was binding between the wt peptide and both importins, but the binding did not induce large changes in the structures of both macromolecules.

The situation was slightly different in the case of the Impα3 and ΔImpα3 complexes with the other mutant peptides. As an example, we described our results with the K65A peptide, and the findings for the other peptides were basically similar to those described here. Where the far-UV CD spectra of the addition and that of the complex with both importins also showed small differences (Figure S6A,B), the fluorescence spectra did not have modifications (either by excitation at 280 or 295 nm) (Figure S6C,D). In general, for the mutant peptides, the changes were smaller than for the wt peptides.

The above experiments were sufficient to conclude that the NLS-NUPR1 peptides interacted with Impα3 or ΔImpα3, but we also carried out ITC experiments to measure the binding affinity. The results (Table 2 and Figure 3) indicate that: (i) the highest affinity towards either Impα3 or ΔImpα3 was that observed for the wt peptide, (ii) the affinity for most of the peptides was higher for binding to ΔImpα3 (the only exceptions were the T68EK69A and pT68 peptides), (iii) removal of Lys65 or Lys69 residues decreased the affinity (and the variations in affinity were higher for ΔImpα3 than for Impα3), and (iv) the phosphorylation or mutation to Glu (phospho-mimics) of Thr68 decreased the affinity by almost one order of magnitude when compared to the other mutations for both importin species. Therefore, the ITC findings mirrored the results obtained by fluorescence: there were lesser structural changes (as reported by fluorescence) in the binding of the peptide mutants than for the wt one, and the affinity of the former peptides for importins was lower (Table 2).

Taking together all these findings, we conclude that the isolated region of NUPR1 comprising its NLS was capable of binding to Impα3 and that this binding was strongly modulated by the phosphorylation state of Thr68 and the charges at positions Lys65 and Lys69.

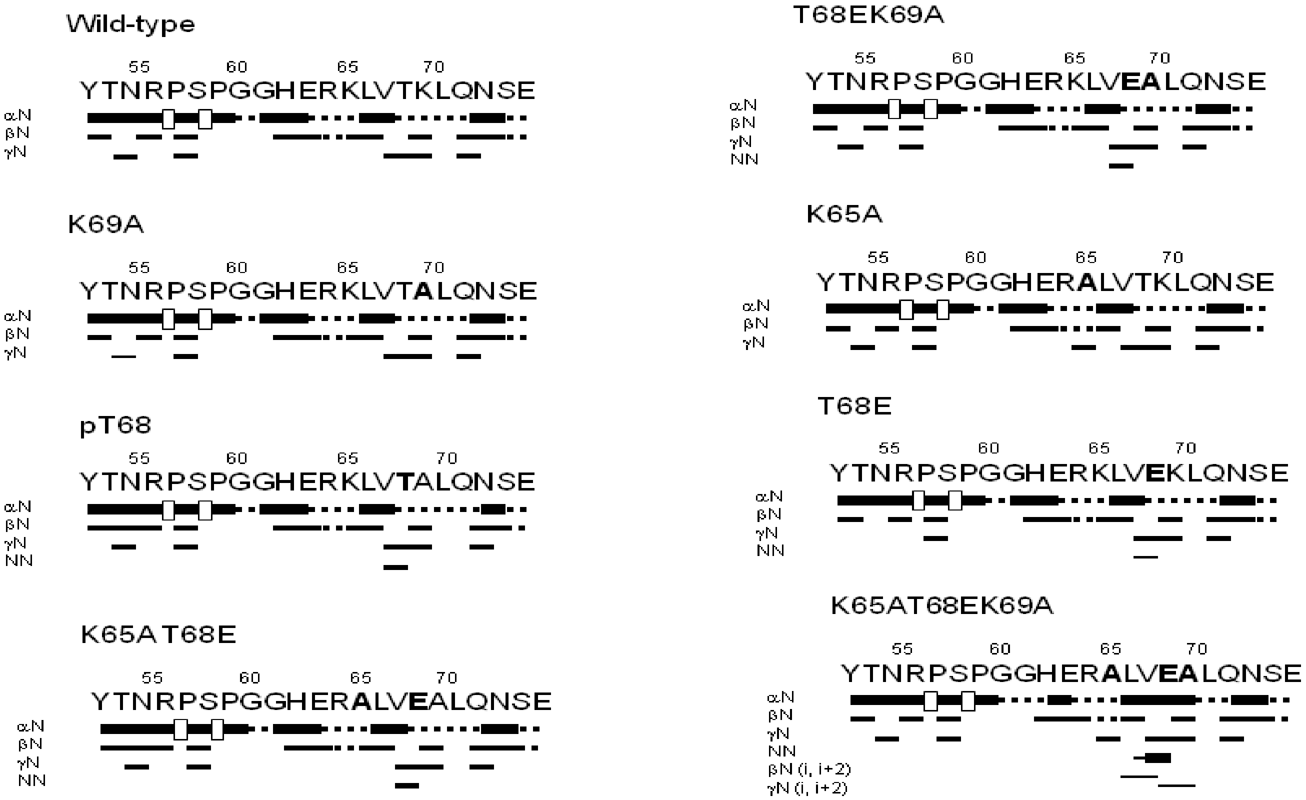

Wild-type
55 60 65 70
YTNRPSPGGHERKLVTKLQNSE
αN
βN
γN
K69A
55 60 65 70
YTNRPSPGGHERKLVTALQNSE
αN
βN
γN
pT68
55 60 65 70
YTNRPSPGGHERKLVTALQNSE
αN
βN
γN
NN
K65AT68E
55 60 65 70
YTNRPSPGGHERALVEALQNSE
αN
βN
γN
NN
T68EK69A
55 60 65 70
YTNRPSPGGHERKLVEALQNSE
αN
βN
γN
NN
K65A
55 60 65 70
YTNRPSPGGHERALVTKLQNSE
αN
βN
γN
T68E
55 60 65 70
YTNRPSPGGHERKLVEKLQNSE
αN
βN
γN
NN
K65AT68EK69A
55 60 65 70
YTNRPSPGGHERALVEALQNSE
αN
βN
γN
NN
βN (i, i+2)
γN (i, i+2)

Figure 1. NMR structural characterization of the nuclear localization sequence (NLS) NUPR1 peptides. Nuclear Overhauser effects (NOEs) are classified into strong, medium, or weak, as represented by the height of the bar underneath the sequence; the signal intensity was judged by visual inspection from the nuclear Overhauser effect spectroscopy (NOESY) experiments. The symbols αN, βN, γN, and NN correspond to the sequential contacts (that is, for instance, the NN corresponds to the NN (i,i + 1) contacts). The corresponding H_α NOEs with the H_δ of the following proline residues are indicated by an open bar in the row corresponding to the αN contacts. The dotted lines indicate NOE contacts that could not be unambiguously assigned due to signal overlap. The numbering of the residues corresponds to that of the whole sequence of NUPR1.

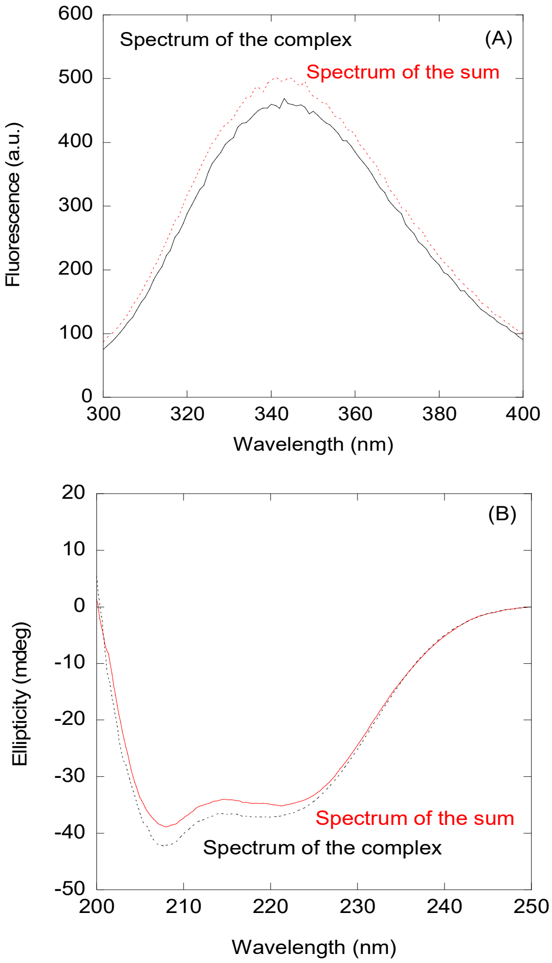


Figure 2. Binding of the wild-type (wt) peptide to importin α3 (Impα3) monitored by spectroscopic techniques: (**A**) Fluorescence spectrum obtained by excitation at 280 nm of the complex between Impα3 and the wt peptide and the addition spectrum obtained by the sum of the spectra of both isolated macromolecules. (**B**) Far-UV CD (ultraviolet circular dichroism) spectrum of the complex between the Impα3 and wt peptides and the additional spectrum obtained by the sum of the spectra of both isolated macromolecules.

3.3. Binding Regions in the Docking of NUPR1 Peptides to Importins

Since we have shown that there was binding between the peptides and both importins, and we have identified the most important residues for attaining such binding, we performed molecular docking to determine details on the location and binding energy of the NUPR1 peptides on the surface of Impα3. When applied to our case, the docking techniques possess three caveats that are worth mentioning explicitly. First, even in the case of our relatively short peptides, the number of degrees of freedom to be considered is too large to be computationally tractable. This number was halved by considering reduced sequences (nine amino acids, corresponding to residues 63–71 of NUPR1), which included all the mutation sites plus at least two more residues at each end. Second, it is impossible with this technique to discriminate differences in the binding between Impα3 and ΔImpα3, and therefore, only the latter protein structure was considered. Third, molecular docking does not take into account the dynamics of a protein-ligand complex, which could also contribute to the binding. Keeping in mind these limitations, the protein surface was blindly explored by considering a volume that included the whole structure and using a high exhaustiveness of search that is equivalent to running multiple (>10) distinct simulations.

Figure 4 summarizes the predictions obtained in our docking calculations. In particular, Figure 4A illustrates the energetically most favorable poses obtained for the wild-type (capped) sequence ERKLVTKLQ mapped on the surface of importin. The best eight poses are reported for clarity and to obtain a more direct comparison with the cluster of the single best pose for each of the eight different peptides (see below, Figure 4D). The results clearly show that the most favorable binding modes cluster into a single location that consists of the major NLS-binding site, located on ARM repeats 2–4. As shown in Figure 4B, the best structure found for our peptide sequence overlaps quite remarkably with that of the NLS of the Epstein-Barr virus EBNA-LP protein, whose structure has been previously determined in crystallography [46]. A number of different amino acids participate in the binding, including some key tryptophan residues (see the details in Figure 4C) that are known to play an important role in the formation of the importin-cargo complex. The binding energy in the docking for the most favorable conformation was −7.2 kcal/mol, indicating a moderate affinity in the low micromolar range. Compared to the experimental values found for the whole wild-type sequence YTNRPSPGGHERALVTKLQNSE (−7.87 and −8.22 kcal/mol for Impα3 and ΔImpα3, respectively; see Table 2), this finding indicates that the reduced docked sequence provides the major contribution to the binding-free energy of the full-length peptide.

Figure 4D shows the best docking poses obtained for the seven mutant sequences compared to the wt one, which is also reported. Again, in this case, all the most favorable binding modes (and, more generally, even the first ten docking poses for each peptide species) clustered in the same location correspond to the major NLS-binding site. This observation suggests that the mutations do not modify essentially the binding location of the peptides but only their affinity towards importin. The calculated binding energies ranged from −5.6 to −6.6 kcal/mol, indicating that any of the explored mutations reduced the binding affinity with respect to the wt sequence, in agreement with our experimental results (Table 2). We observed a poor correlation between the computational and experimental rankings of the mutated peptides in terms of affinity towards the protein, although this could reasonably be explained, because the experimental binding energies are, in most cases, very close to each other (Table 2). This finding did not let us push too far the interpretation of our results in terms of the molecular details that assist the binding. Nevertheless, the contribution of the protein tryptophan residues to the binding still seemed to be, in all cases, an important determinant (even though we did not observe changes in the fluorescence spectra (either by excitation at 280 or 295 nm) when binding for some of the mutant peptide sequences was explored, Figure S6).

To sum up, a number of important conclusions can be drawn from the docking results reported: (i) all the sequences investigated interacted with the same region of importin; (ii) this region matched unambiguously with the major NLS-binding site of the protein; (iii) the ligand with the highest binding affinity corresponded to the wt sequence of NUPR1 (in agreement with the experimental results from ITC; Table 2); (iv) the major contribution to the binding energy of the parent peptides (i.e., those used in this work) was due to such a restricted sequence portion, which includes only nine residues (and this region includes Lys65, Thr68, and Lys69); (v) this essential sequence fragment corresponded to the core region of the predicted NLS of NUPR1; (vi) the binding region roughly mapped around Thr68 (where the residue name and number refers to wild-type, intact NUPR1 numbering), which therefore appears to be a key amino acid; and (vii) the most favorable predicted structure for the NLS region of wild-type NUPR1 essentially overlapped with the conformation of the NLS of a different protein (the Epstein-Barr virus EBNA-LP protein) determined in crystallography.

Table 2. Thermodynamic parameters at 298 K in the binding reaction of NLS NUPR1 peptides to the two importin species [a].

	Impα3				**ΔImpα3**			
Peptide	**K_d (μM)**	**ΔH (kcal/mol)**	**$-T\Delta S$ (kcal/mol)**	**n**	**K_d (μM)**	**ΔH (kcal/mol)**	**$-T\Delta S$ (kcal/mol)**	**n**
wt	1.7	0.8	−8.7	0.9	0.95	−3.7	−4.5	1.0
K65A	3.9	−2.8	−4.6	1.4	2.7	−10.2	2.6	1.4
K69A	11	−10.8	4.0	1.3	7.6	−21.3	14.3	1.4
T68E	22	−11.1	4.7	(1)	12	−17.5	10.8	(1)
K65AT68E	21	−7.8	1.4	(1)	14	−17.9	11.3	(1)
T68EK69A	17	−7.5	1.0	(1)	17	−21.2	14.7	(1)
K65AT68EK69A	27	−16.3	9.1	(1)	24	−28.5	22.2	(1)
pT68	27	−14.8	3.6	(1)	29	−28.2	22.0	(1)

[a] Relative error in K_d (dissociation constant) is 30%, absolute errors in ΔH (enthalpy) and $-T\Delta S$ (entropy) are 0.5 and 0.7 kcal/mol, respectively, and absolute error in n (the stoichiometry) is 0.2. The parenthesis in n values indicate that this parameter had to be fixed in order to get convergence in the fit due to low affinity.

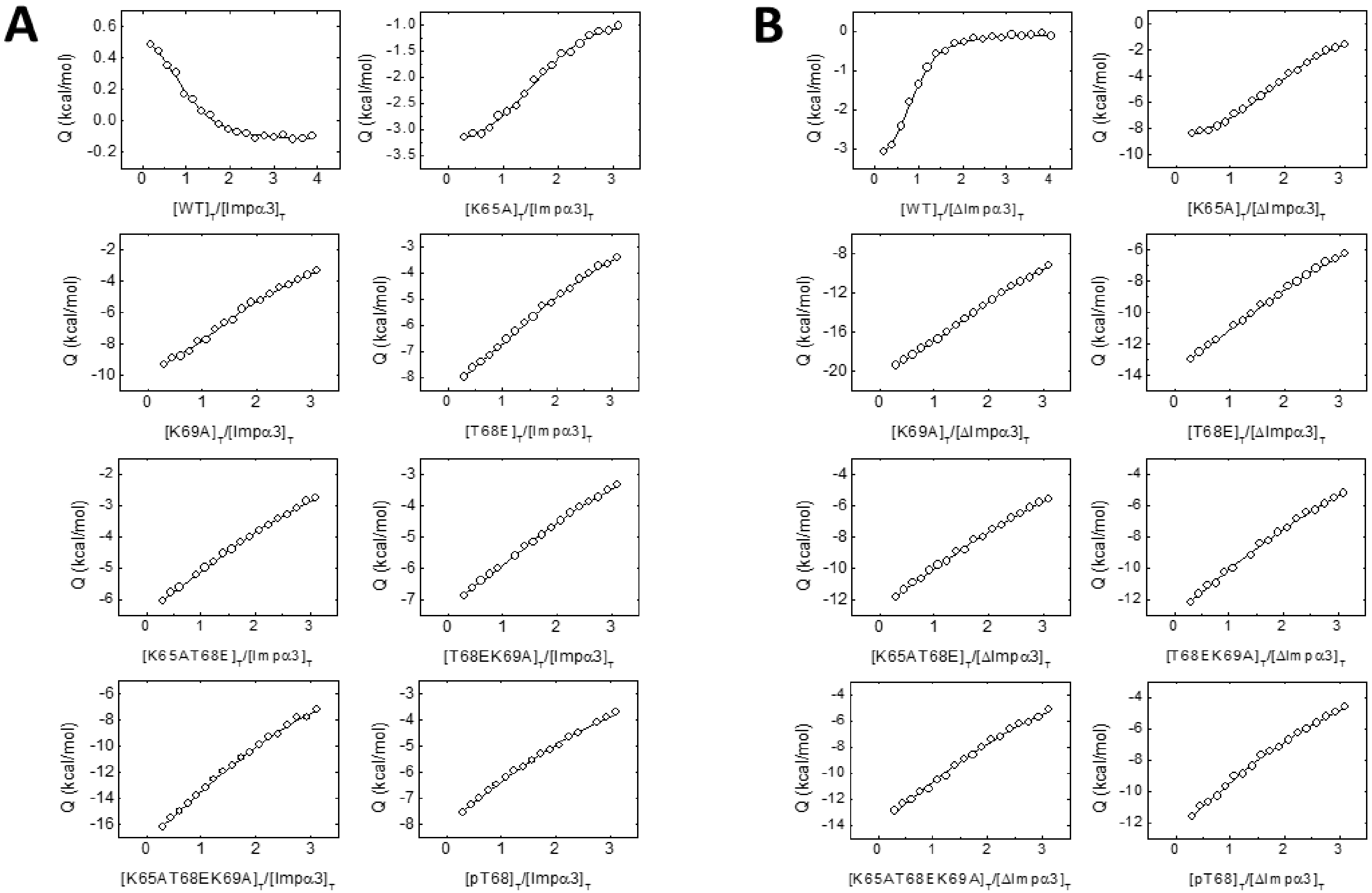


Figure 3. The interaction of the wt and mutant NLS NUPR1 peptides with both importins as measured by isothermal titration calorimetry (ITC). Interaction isotherms (ligand normalized heat effect per injection as a function of the ligand:protein molar ratio) with Impα3 (**A**) and ΔImpα3 (**B**) are shown. Binding parameters were estimated by a nonlinear least-squares regression data analysis of the interaction isotherms applying a single ligand binding site model, implemented in Origin 7.0.

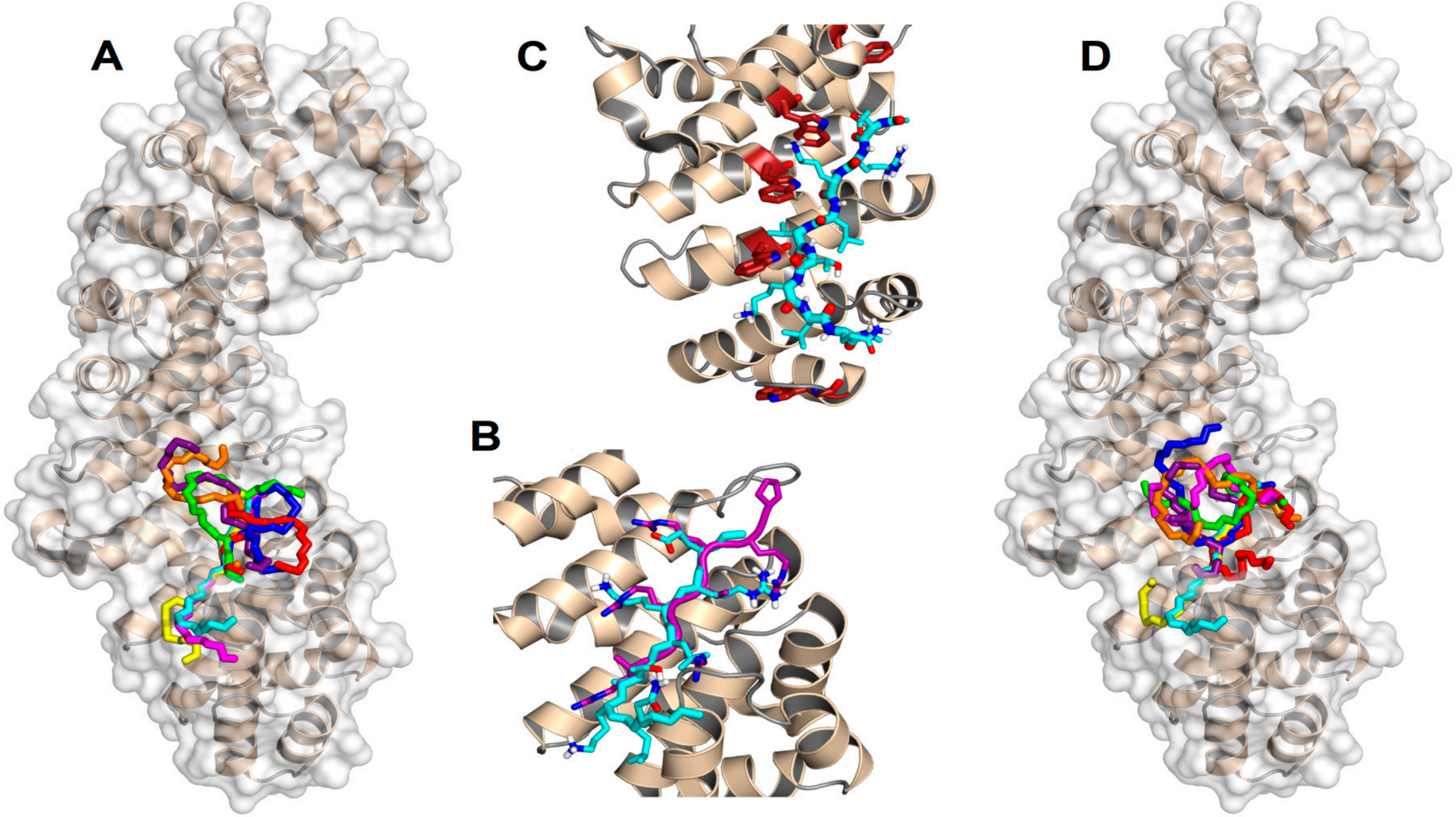


Figure 4. Predicted docking poses for the NLS of NUPR1 on importin. (**A**) Backbone (–N–C$^{\alpha}$–C– atoms) representation of the best eight docking poses on ΔImportin for the wt sequence ERKLVTKLQ (the N terminus is on the top), which constitutes the core region for the NLS of NUPR1. (**B**) Most favorable binding pose for the same sequence (cyan), compared to the crystallographic conformation [46] of the NLS of the Epstein-Barr virus EBNA-LP protein (purple). For clarity, atoms are shown in standard colors only in the side chains of the two peptides, and the main-chain O and H atoms are omitted; apolar H atoms are not present. (**C**) Trp residues (brown) in the major NLS-binding site of importin play a key role in the binding of the most favorable conformation of the NLS of wild-type NUPR1. The view is slightly rotated with respect to previous representations to evidence the tryptophan side chains. (**D**) Most favorable docking poses for the eight peptide sequences: wild type (cyan), K65A (magenta), K69A (yellow), T68E (blue), K65AT68E (red), T68EK69A (green), K65AT68EK69A (orange), and pT68 (violet). PyMol was used for all displays.

4. Discussion

4.1. Molecular Mechanisms for Impα3 Recognition of NUPR1: The Influence of Lys65 and Lys69

In this work, we tried first to find out whether the theoretically predicted NLS region of NUPR1 was capable of binding in isolation to Impα3. Second, we tried to elucidate, for the first time, the molecular bases behind the binding of an NLS region of an IDP to an importin. Our results indicate that the isolated, wild-type NLS region of NUPR1 interacted with the intact Impα3 and ΔImpα3, with an affinity similar to that for intact NUPR1 (1.4 μM, Figure S4), and within the same range measured for the affinities to natural partners of NUPR1 [22,24,51] and synthetic molecules [23,52]. Furthermore, our results also address the molecular importance of IBB in the binding of cargos to importins.

As it happens for the intact NUPR1 (whose dissociation constants are 1.4 μM for Impα3 and 0.44 μM for ΔImpα3 (Figure S4)), the wt peptide bound to ΔImpα3 with a two-fold larger affinity (0.95 μM) than that for Impα3 (Table 2) (1.7 μM). These findings allow us to draw several conclusions. First, the presence of the IBB region (which contains a large quantity of lysine amino acids) exerts an autoinhibitory effect, and the domain hampers the entrance of the NLS peptide into the major NLS-binding region of Impα3, as it has been suggested in other studies with well-folded proteins [9]. However, this is the first time such a hypothesis is tested in an IDP. Modulation of the assembly complex formation between importins and their cargos has been attributed to the IBB domain [4]; this domain has been found to be involved even in the formation of a homodimeric species between importins [53], with a reduced ability to bind cargos. Second, although the affinities of the wt peptide for both importins were smaller than those for intact NUPR1, many of the interactions implicated in the binding to importin could be ascribed to a region comprised within the wt peptide, as concluded from the similarities among the dissociation constants (0.44 (intact NUPR1) and 0.95 μM (wt peptide) for ΔImpα3 and 1.4 (intact NUPR1) and 1.7 μM (wt peptide) for Impα3). Third, given the similarities among the affinity constants for Impα3 of the wt peptide and NUPR1, the peptide could be used as a lead compound to design an inhibitor of its nuclear translocation.

We have previously shown in vivo that a mutant of NUPR1 at positions Lys65, Lys69, Lys76, and Lys77 is present through the whole cell, whereas the wild-type NUPR1 species is localized exclusively into the nucleus [25]. In this work, we have found that the mutation Lys65Ala decreased two-fold the affinity for Impα3, and the mutation Lys69Ala decreased six-fold the affinity. Thus, the decrease in the affinity was larger with the removal of Lys69, probably indicating that this residue makes more contacts with importin, as pinpointed by our docking models. In fact, we observed in the simulation that both lysine residues were involved in hydrophobic and polar contacts (the latter with their NH_3^+ moieties), with residues of importin α. The removal of the long side chains would disfavor those contacts, thus decreasing the affinity (Table 2). The importance of lysines is key in determining the binding to importins of other well-folded proteins through their disordered NLS regions, as shown by several structural studies [3,8,54,55]. It could be thought that our study does not provide new mechanistic insight into the function of importins, because the results obtained with an IDP pinpoint, for the first time, the importance of positive charges (as it happens in folded proteins) in the binding of their cargos; however, to the best of our knowledge, this is the first reported case where the importance of such residues is addressed in vitro for the NLS of an IDP, and our results acquire more relevance considering recent findings, where it has been suggested that IDPs do not require the presence of importins to be translocated into the nucleus, although demonstrated mostly for acidic proteins [56]. Then, our results indicated that IDPs require the help of importins to be translocated into the nucleus, and it seems that the rules governing such processes are similar to those observed in well-folded proteins.

The same decrease in affinity was observed for the K65A and K69A peptides towards ΔImpα3, but, compared to the wt peptide, the variation was larger than that observed for both mutants with Impα3 (Table 2). Furthermore, as it happens with the intact importin, the decrease in affinity was larger for the K69A peptide. These findings indicate that, although the IBB region maintains its independence within the whole Impα3 in terms of conformation, its removal may either alter the structure of some regions of the ARM repeats involved in the major NLS-binding site (which relies on hydrophobic contacts to anchor the cargo, therefore altering its docking) or, alternatively, IBB removal may change the whole protein dynamics and its stability.

4.2. Molecular Mechanisms for Impα3 Recognition of NUPR1: The Influence of Thr68 and Its Phosphorylation-Triggered Conformational Switch

Apart from the importance of the two lysines of NUPR1 in the binding to importins, we also wanted to address the importance of Thr68. It is well-established that Thr68 is a key residue in the binding of NUPR1 to any partner, either natural or synthetic [22–24,57]; in fact, together with Ala33, it constitutes one of the two hotspot regions of NUPR1. We decided to address such a question by following two approaches: (i) we mutated Thr68 to Glu to have the phospho-mimics, and (ii) we synthesized a peptide with the phosphorylated Thr (pT68 peptide).

Among all the mutants explored, the peptide with the smallest affinity for Impα3 or ΔImpα3 (~30 μM for both species) was the pThr68 peptide (Table 2). Phosphorylation affects the binding probably by inhibiting long-range electrostatic contacts with both importins. Where the affinity of the wt peptide for ΔImpα3 was larger, the changes due to the addition of the phosphate group in the pThr68 peptide were even larger, further pinpointing subtle structural changes in the major NLS-binding region upon removal of the IBB. Around a third of the eukaryotic proteins can be phosphorylated, and the majority of those phosphorylation sites belong to intrinsically disordered regions because of their accessibility to kinases [14]. Phosphorylation is a key regulatory mechanism in translation, transcription, and other processes.

The phospho-mimetic peptide of NLS-NUPR1, the T68E peptide, also showed a smaller affinity for both importins than the wt one (22 μM for Impα3 and 12 μM for ΔImpα3, Table 2), but the decrease was not as large as that in the pT68 peptide (27 μM for Impα3 and 29 μM for ΔImpα3, Table 2), indicating that the phospho-mimics did not cause the same effect as phosphorylation. Phosphorylation at Thr68 replaces the neutral OH (hydroxyl) group with a tetrahedral PO_4^{2-} (phosphoryl group) with two negative charges, which modifies the electrostatic, chemical, and steric properties of the threonine environment. The double-negative charge of the PO_4^{2-} and its large surrounding hydration shell make the situation chemically different from the Glu phospho-mimic, which has a smaller hydration shell and a single negative charge. Differences among the affinities of phospho-mimics and phosphorylated threonines for a well-folded protein have been also observed in the affinities measured in other protein systems [58], as well as in other IDPs [59].

The values of the affinity constants of the K65AT68E and T68EK69A peptides were similar to that of the T68E peptide (Table 2). This finding indicates that: (i) the effect of Thr68Glu in the binding to importins surpassed those caused by substitutions of the single lysines (and then, Thr68 must have a greater importance in the interaction), and (ii) the effect of removing a lysine when the threonine is phospho-mimicked is not additive for the double mutants, probably because the remnant lysine establishes electrostatic interactions with the glutamic residue. However, the accumulation of the three mutations (in the K65AT68EK69A peptide) led to a large decrease of the affinity constant (Table 2), further highlighting the influence of electrostatic effects between the lysines and the phospho-mimics in the bindings with the two importins. Other studies of phosphorylation of threonines in IDPs indicate that the proximity of arginines can stabilize the charge of the phosphoryl moiety and the stabilization of turn-like structures [60]. We suggest that, in the case of NUPR1, lysines, instead of arginines, would play the role of stabilizing the conformation.

Interestingly enough, the peptides containing the phospho-mimic mutation (T68E) or the phosphorylated Thr68 (pT68 peptide) did show an NN(i,i + 1) NOE (Figure 1) between Val67 and phosphorylated Thr68 (or Glu68). We did not observe such a NOE in the intact NUPR1 when we assigned it [22]. It could be thought that the absence of such a NOE in the wt peptide may be due to the fact that the chemical shifts of the amide protons of those residues (Val67 and Thr68) were similar (Table S1), and then, the NOE could not be observed because of its proximity to the spectrum diagonal. However, the chemical shifts of amides of both residues in the K65A peptide were different enough (8.25 and 8.35 ppm for Val67 and Thr68, respectively; Table S6) to allow for its detection, and nevertheless, we did not observe any NOE (Figure S3). Thus, the presence of such a NOE, although it is weak in intensity, indicates that, upon phosphorylation, the two residues populated a turn-like conformation [40]; the presence of this turn is further supported by the observation of $\beta N(i,i+2)$ and $\gamma N(i,i+2)$ NOEs and an additional $NN(i,i+1)$ contact for the K65AT68EK69A peptide involving residues Leu66–Thr68 (Figure 1), due to the large, intrinsic propensity of alanine to populate helix-like conformations [61,62]. As the affinity of the peptides for both importins decreased when Thr68 was phosphorylated or was phospho-mimicked (Table 2), we can conclude that the decrease in the affinity of peptides upon phosphorylation was structurally related to a conformational switch around Thr68, as a consequence of the introduced negative charge, shifting the population at equilibrium from a random-coil conformation to a turn-like one. The decrease in affinity for both importins may be related to the reduction in entropy of the polypeptide chain upon acquisition of the turn-like conformation and a concomitant conformational energetic penalty for the binding. Interestingly enough, two decades ago, we showed by using FTIR (Fourier transform infra-red spectroscopy) and CD that the unspecific phosphorylation of the serines and threonines in NUPR1 led to a higher population of α-helix- and/or turn-like conformations in the intact protein [21]; at the moment, however, we do not have any evidence for the biological importance of the particular phosphorylation of Thr68 in vivo. Nonetheless, we have recently shown that the mutation of Thr68 to Gln hampers the formation of several complexes of NUPR1 with other proteins involved in SUMOylation processes [50]. Our previous result is confirmed in this work by our new findings obtained with Thr68. Phosphorylation, as well as other post-translational modifications, can affect protein conformations: (i) on a local scale—for instance by affecting the population of *cis* proline isomers [63], (ii) determining a change of entropy of the conformational ensemble [64], (iii) modulating the binding to other macromolecules and triggering phase separation [65], (iv) in an allosteric manner, by affecting distant residues from the phosphorylation site [66], and, (iv) causing a conformational change [67,68]. Conformational switching affecting a threonine in several IDPs has been described [59,69,70]. For instance, the phosphorylation of Thr51 in the IDP prostate-associated gene protein increases the population of transient turn-like populations [70]; the difference with our results is that the turn-like structures in NUPR1 were stabilized in a much shorter polypeptide region, although we cannot rule out that phosphorylation at other sites of NUPR1 could help in stabilizing this conformation. On the other hand, the p27 protein, which modulates the mammalian cell cycle by the inhibition of cyclin-dependent kinases, contains some disordered regions, and the phosphorylation of residue Thr157 in breast cancer cells prevents its interaction with the nuclear import machinery, leading to the accumulation of this protein in the cytoplasm, whereas it is normally found in the nucleus [69]; however, no indication on the particular structure acquired upon phosphorylation at Thr157 has been provided. Finally, it is important to note that recent theoretical molecular dynamic simulations have shown that the binding of importin α to heterochromatin protein 1 α is modulated by phosphorylation at residues in its importin-binding region [71].

Thr68 is, together with the polypeptide patch around Ala33, the hotspot region of NUPR1, involved in binding to its natural partners [21,22,24,51] and to other synthetic molecules and macromolecules [23,52,57]. We have previously observed that the mutation of Thr68 to glutamine hampers the binding to those other molecules [24,51]. Such a mutation will probably cause a shift of the ensemble population from a random-coil towards turn-like conformations, and it is the adoption of such a local fold that hampers bindings to those other natural partners or synthetic molecules. Moreover, as the affinity of NUPR1 to its partners is basically the same in all cases described to date [21,22,24,51,52,57], its binding features can also be modulated by phosphorylation at Thr68 at least partially, since the region around Ala33 is also involved in the binding. In addition, since this threonine is also associated with the binding of drugs strongly effective against pancreatic cancer in mice [23,32], we hypothesize that the molecular effects of such drugs could be the induction of a stable fold (turn-like) by this polypeptide region, besides competitive steric hindrance, preventing binding to other natural partners of NUPR1, and hampering the protein cascades where it is involved.

5. Conclusions

We have described the interaction between the NLS region of NUPR1, a nuclear intrinsically disordered protein involved in cancer, and Impα3 by using a series of peptides comprising that polypeptide patch. Binding to Impα3 is modulated by the charges of Lys64 and Lys69 but, most importantly, by phosphorylation at Thr68, which constitutes an entropy-driven conformational switch, shifting the population of the dynamic ensemble towards a turn-like conformation. As Thr68 is also a hotspot for NUPR1 interactions, these results open the venue to modulating the binding to its partners by targeting this residue. Furthermore, it also suggests a possible mechanism for the action of drugs targeting NUPR1, which also bind through Thr68.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/9/1313/s1, Figure S1: Structural features of NLS-NUPR1 peptides as monitored by far-UV CD, Figure S2: Structural features of NLS-NUPR1 peptides as monitored by 1D-^{1}H-NMR, Figure S3: The amide region of 2D-^{1}H-NOESY spectra of NLS-NUPR1 peptides, Figure S4: Interaction between (left) Impα3 and (right) ΔImp α 3 with full-length NUPR1 as observed by ITC, Figure S5: Thermal denaturation of the complexes followed by spectroscopic techniques, Figure S6: Interaction between Imp α3 and ΔImp α3 with K65A peptide measured by different spectroscopic techniques, Table S1: Chemical shifts (δ, ppm from TSP) of wt peptide in aqueous solution (pH 7.2, 283K), Table S2: Chemical shifts (δ, ppm from TSP) of K69A peptide in aqueous solution (pH 7.2, 283 K), Table S3: Chemical shifts (δ, ppm from TSP) of pT68 peptide in aqueous solution (pH 7.2, 283 K), Table S4: Chemical shifts (δ, ppm from TSP) of T68EK69A peptide in aqueous solution (pH 7.2, 283 K), Table S5: Chemical shifts (δ, ppm from TSP) of K65AT68E peptide in aqueous solution (pH 7.2, 283 K), Table S6: Chemical shifts (δ, ppm from TSP) of K65A peptide in aqueous solution (pH 7.2, 283 K), Table S7: Chemical shifts (δ, ppm from TSP) of T68E peptide in aqueous solution (pH 7.2, 283 K), Table S8: Chemical shifts (δ, ppm from TSP) of K65AT68EK69A peptide in aqueous solution (pH 7.2, 283 K).

Author Contributions: Conceptualization, J.L.N., B.R., A.V.-C., O.A., and J.L.I.; methodology, J.L.N., B.R., A.V.-C., O.A., and J.L.I.; investigation, J.L.N., B.R., A.V.-C., M.P.-S., and A.J.-A.; data analysis, J.L.N., B.R., A.V.-C., and A.J.-A.; writing—original draft preparation, J.L.N., B.R., and A.V.-C.; writing—review and editing, J.L.N., B.R., J.L.I., M.P.-S., A.J.-A., O.A., and A.V.-C.; and funding acquisition, J.L.N., A.V.-C., O.A., and J.L.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Ministry of Economy and Competitiveness and European ERDF Funds (MCIU/AEI/FEDER, EU) (RTI2018-097991-B-I00 to J.L.N. and BFU2016-78232-P to A.V.C.); La Ligue Contre le Cancer, INCa, Canceropole PACA, and INSERM to J.L.I.; Miguel Servet Program from Instituto de Salud Carlos III (CPII13/00017 to OA); Fondo de Investigaciones Sanitarias from Instituto de Salud Carlos III and European Union (ERDF/ESF, "Investing in your future") (PI15/00663 and PI18/00349 to O.A.); Diputación General de Aragón (Protein Targets and Bioactive Compounds Group E45_17R to A.V.C. and Digestive Pathology Group B25_17R to O.A.); and the Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd). The NMR equipment used in this work was funded by the Generalitat Valenciana and cofinanced with ERDF funds (OP ERDF of Comunitat Valenciana 2014-2020).

Acknowledgments: We thank J. K. Forwood (Charles Sturt University, Waga Waga, Autralia) for the kind gift of the ΔImpα3 vector. B.R. acknowledges the kind hospitality and use of computational resources in the European Magnetic Resonance Center (CERM), Sesto Fiorentino (Florence), Italy. We thank the three anonymous reviewers for their helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Stewart, M. Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell Boil.* **2007**, *8*, 195–208. [CrossRef]
2. Bednenko, J.; Cingolari, G.; Gerace, L. Nucleo-cytoplasmic transport navigating the channel. *Traffic* **2003**, *4*, 127–135. [CrossRef] [PubMed]
3. Cingolani, G.; Bednenko, J.; Gillespie, M.T.; Gerace, L. Molecular basis for the recognition of a non-classical nuclear localization signal by importin beta. *Mol. Cell* **2002**, *10*, 1345–1353. [CrossRef]

4. Goldfarb, D.S.; Corbett, A.H.; Mason, D.A.; Harreman, M.T.; Adam, S.A. Importin α: A multipurpose nuclear-transport receptor. *Trends Cell Boil.* **2004**, *14*, 505–514. [CrossRef] [PubMed]
5. Pumroy, R.A.; Cingolani, G. Diversification of importin-α isoforms in cellular trafficking and disease states. *Biochem. J.* **2015**, *466*, 13–28. [CrossRef]
6. Mason, D.A.; Stage, D.E.; Goldfarb, D. Evolution of the metazoan-specific importin α gene family. *J. Mol. Evol.* **2009**, *68*, 351–365. [CrossRef]
7. Miyamoto, Y.; Loveland, K.L.; Yoneda, Y. Nuclear importin α and its physiological importance. *Commun. Integr. Boil.* **2012**, *5*, 220–222. [CrossRef]
8. Smith, K.M.; Tsimbalyuk, S.; Edwards, M.R.; Cross, E.M.; Batra, J.; Da Costa, T.P.S.; Aragão, D.; Basler, C.; Forwood, J. Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses. *Nat. Commun.* **2018**, *9*, 3703. [CrossRef]
9. Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α. *Nat. Struct. Biol.* **1999**, *6*, 388–397. [CrossRef]
10. Berlow, R.B.; Dyson, H.J.; Wright, P.E. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J. Mol. Boil.* **2018**, *430*, 2309–2320. [CrossRef]
11. Xie, H.M.; Vucetic, S.; Iakoucheva, L.M.; Oldfield, C.J.; Dunker, A.K.; Uversky, V.N.; Obradovic†, Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **2007**, *6*, 1882–1898. [CrossRef] [PubMed]
12. Babu, M.M.; Van Der Lee, R.; De Groot, N.S.; Gsponer, J. Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Boil.* **2011**, *21*, 432–440. [CrossRef] [PubMed]
13. Gsponer, J.; Futschik, M.E.; Teichmann, S.A.; Babu, M.M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science* **2008**, *322*, 1365–1368. [CrossRef]
14. Bah, A.; Forman-Kay, J.D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Boil. Chem.* **2016**, *291*, 6696–6705. [CrossRef]
15. Launay, H.; Receveur-Bréchot, V.; Carrière, F.; Gontero, B. Orchestration of algal metabolism by protein disorder. *Arch. Biochem. Biophys.* **2019**, *672*, 108070. [CrossRef]
16. Mallo, G.V.; Fiedler, F.; Calvo, E.L.; Ortiz, E.M.; Vasseur, S.; Keim, V.; Morisset, J.; Iovanna, J.L. Cloning and Expression of the Rat p8 cDNA, a New Gene Activated in Pancreas during the Acute Phase of Pancreatitis, Pancreatic Development, and Regeneration, and Which Promotes Cellular Growth. *J. Boil. Chem.* **1997**, *272*, 32360–32369. [CrossRef]
17. Chowdhury, U.R.; Samant, R.S.; Fodstad, O.; Shevde, L.A. Emerging role of nuclear protein 1 (NUPR1) in cancer biology. *Cancer Metastasis Rev.* **2009**, *28*, 225–232. [CrossRef] [PubMed]
18. Goruppi, S.; Iovanna, J.L. Stress-inducible Protein p8 Is Involved in Several Physiological and Pathological Processes. *J. Boil. Chem.* **2009**, *285*, 1577–1581. [CrossRef]
19. Cano, C.; Hamidi, T.; Sandi, M.J.; Iovanna, J.L. Nupr1: The Swiss-knife of cancer. *J. Cell. Physiol.* **2010**, *226*, 1439–1443. [CrossRef] [PubMed]
20. Malicet, C.; Giroux, V.; Vasseur, S.; Dagorn, J.C.; Neira, J.L.; Iovanna, J.L. Regulation of apoptosis by the p8/prothymosin alpha complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2671–2676. [CrossRef]
21. Encinar, J.A.; Mallo, G.V.; Mizyrycki, C.; Giono, L.E.; González-Ros, J.M.; Rico, M.; Cánepa, E.T.; Moreno, S.; Neira, J.L.; Iovanna, J.L. Human p8 is a HMG-I/Y-like protein with DNA binding activity enhanced by phosphorylation. *J. Boil. Chem.* **2000**, *276*, 2742–2751. [CrossRef]
22. Aguado-Llera, D.; Hamidi, T.; Doménech, R.; Pantoja-Uceda, D.; Gironella, M.; Santoro, J.; Velázquez-Campoy, A.; Neira, J.L.; Iovanna, J.L. Deciphering the binding between Nupr1 and MSL1 and Their DNA-Repairing Activity. *PLoS ONE* **2013**, *8*, e78101. [CrossRef] [PubMed]
23. Neira, J.L.; Bintz, J.; Arruebo, M.; Rizzuti, B.; Bonacci, T.; Vega, S.; Lanas, A.; Velázquez-Campoy, A.; Iovanna, J.L.; Abián, O. Identification of a Drug Targeting an intrinsically disordered protein involved in pancreatic adenocarcinoma. *Sci. Rep.* **2017**, *7*, 39732. [CrossRef] [PubMed]
24. Santofimia-Castaño, P.; Rizzuti, B.; Pey, A.L.; Soubeyran, P.; Vidal, M.; Urrutia, R.; Iovanna, J.L.; Neira, J.L. Intrinsically disordered chromatin protein NUPR1 binds to the C-terminal region of Polycomb RING1B. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6332–E6341. [CrossRef] [PubMed]
25. Valacco, M.P.; Varone, C.L.; Malicet, C.; Cánepa, E.T.; Iovanna, J.L.; Moreno, S. Cell growth-dependent subcellular localization of p8. *J. Cell. Biochem.* **2006**, *97*, 1066–1079. [CrossRef]

26. Jäkel, S.; Mingot, J.-M.; Schwarzmaier, P.; Hartmann, E.; Görlich, D. Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO J.* **2002**, *21*, 377–386. [CrossRef]
27. Marvaldi, L.; Panayotis, N.; Alber, S.; Dagan, S.Y.; Okladnikov, N.; Koppel, I.; Di Pizio, A.; Song, D.-A.; Tzur, Y.; Terenzio, M.; et al. Importin α3 regulates chronic pain pathways in peripheral sensory neurons. *Science* **2020**, *369*, 842–846. [CrossRef]
28. Gill, S.C.; Von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319–326. [CrossRef]
29. Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ(1-40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magn. Reson. Chem.* **2002**, *40*, S89–S97. [CrossRef]
30. Neira, J.L.; Hornos, F.; Bacarizo, J.; Camara-Artigas, A.; Gómez, J. The monomeric species of the regulatory domain of Tyrosine Hydroxylase has a low conformational stability. *Biochemistry* **2016**, *55*, 3418–3431. [CrossRef]
31. Benjwal, S.; Verma, S.; Röhm, K.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* **2006**, *15*, 635–639. [CrossRef] [PubMed]
32. Santofimia-Castaño, P.; Xia, Y.; Lan, W.; Zhou, Z.; Huang, C.; Peng, L.; Soubeyran, P.; Velázquez-Campoy, A.; Abian, O.; Rizzuti, B.; et al. Ligand-based design identifies a potent NUPR1 inhibitor exerting anticancer activity via necroptosis. *J. Clin. Investig.* **2019**, *129*, 2500–2513. [CrossRef] [PubMed]
33. Cavanagh, J.; Fairbrother, W.J.; Palmer, A.G.; Skelton, N.J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, NY, USA, 1996.
34. Wilkins, D.K.; Grimshaw, S.B.; Receveur, V.; Dobson, C.M.; Jones, J.A.; Smith, L.J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, *38*, 16424–16431. [CrossRef] [PubMed]
35. Marion, D.; Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967–974. [CrossRef]
36. Bax, A.; Davis, D.G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985**, *65*, 355–360. [CrossRef]
37. Kumar, A.; Ernst, R.; Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6. [CrossRef]
38. Cavanagh, J.; Rance, M. Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. *J. Magn. Reson.* **1992**, *96*, 670–678. [CrossRef]
39. Piotto, M.; Saudek, V.; Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992**, *2*, 661–665. [CrossRef]
40. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, NY, USA, 1986.
41. Kjaergaard, M.; Brander, S.; Poulsen, F.M. Random coil chemical shift for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* **2011**, *49*, 139–149. [CrossRef]
42. Kjaergaard, M.; Poulsen, F.M. Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **2011**, *50*, 157–165. [CrossRef]
43. Bienkiewicz, E.A.; Lumb, K.J. Random-coil chemical shifts of phosphorylated amino acids. *J. Biomol. NMR* **1999**, *15*, 203–206. [CrossRef] [PubMed]
44. Hendus-Altenburger, R.; Fernandes, C.B.; Bugge, K.; Kunze, M.B.A.; Boomsma, W.; Kragelund, B.B. Random coil chemical shifts for serine, threonine and tyrosine phosphorylation over a broad pH range. *J. Biomol. NMR* **2019**, *73*, 713–725. [CrossRef] [PubMed]
45. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461. [CrossRef] [PubMed]
46. Nakada, R.; Matsuura, Y. Crystal structure of importin–α bound to the nuclear localization signal of Epstein–Barr virus EBNA–LP protein. *Protein Sci.* **2017**, *26*, 1231–1235. [CrossRef] [PubMed]

47. Forli, S.; Huey, R.; Pique, M.E.; Sanner, M.F.; Goodsell, D.S.; Olson, A.J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016**, *11*, 905–919. [CrossRef]
48. Grande, F.; Rizzuti, B.; Occhiuzzi, M.A.; Ioele, G.; Casacchia, T.; Gelmini, F.; Guzzi, R.; Garofalo, A.; Statti, G. Identification by molecular docking of homoisoflavones from *Leopoldia comosa* as ligands of estrogen receptors. *Molecules* **2018**, *23*, 894. [CrossRef]
49. Grimsley, G.R.; Huyghues-Despointes, B.M.; Pace, C.N.; Scholtz, J.M. Measuring the Conformational Stability of a Protein by NMR. *Cold Spring Harb. Protoc.* **2006**, *2006*, 253–259. [CrossRef]
50. Lan, W.; Santofimia-Castaño, P.; Swayden, M.; Xia, Y.; Zhou, Z.; Audebert, S.; Camoin, L.; Huang, C.; Peng, L.; Jiménez-Alesanco, A.; et al. ZZW-115-dependent inhibition of NUPR1 nuclear translocation sensitizes cancer cells to genotoxic agents. *JCI Insight* **2020**, 138117. [CrossRef]
51. Neira, J.L.; López, M.B.; Sevilla, P.; Rizzuti, B.; Camara-Artigas, A.; Vidal, M.; Iovanna, J.L. The chromatin nuclear protein NUPR1L is intrinsically disordered and binds to the same proteins as its paralogue. *Biochem. J.* **2018**, *475*, 2271–2291. [CrossRef]
52. Santofimia-Castaño, P.; Rizzuti, B.; Abian, O.; Velázquez-Campoy, A.; Iovanna, J.L.; Neira, J.L. Amphipathic helical peptides hamper protein-protein interactions of the intrinsically disordered chromatin nuclear protein 1 (NUPR1). *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862*, 1283–1295. [CrossRef]
53. Miyatake, H.; Sanjoh, A.; Unzai, S.; Matsuda, G.; Tatsumi, Y.; Miyamoto, Y.; Dohmae, N.; Aida, Y. Crystal structure of human Importin-α1 (Rch1), revealing a potential autoinhibition mode involving homodimerization. *PLoS ONE* **2015**, *10*, e0115995. [CrossRef] [PubMed]
54. Sankhala, R.S.; Lokareddy, R.K.; Begum, S.; Pumroy, R.A.; Gillilan, R.E.; Cingolani, G. Three-dimensional context rather than NLS amino acid sequence determines importin α subtype specificity for RCC1. *Nat. Commun.* **2017**, *8*, 979. [CrossRef] [PubMed]
55. Pumroy, R.A.; Ke, S.; Hart, D.J.; Zacharie, U.; Cingolani, G. Molecular determinants for nuclear import if influenza A PB2 by importin alpha isoforms 3 and 7. *Structure* **2015**, *23*, 374–384. [CrossRef] [PubMed]
56. Junod, S.L.; Kelich, J.M.; Ma, J.; Yang, W. Nucleocytoplasmic transport of intrinsically disordered proteins studied by high–speed super–resolution microscopy. *Protein Sci.* **2020**, *29*, 1459–1472. [CrossRef]
57. Neira, J.L.; Correa, J.; Rizzuti, B.; Santofimia-Castaño, P.; Abián, O.; Velázquez-Campoy, A.; Fernandez-Megia, E.; Iovanna, J.L. Dendrimers as competitors of protein–protein interactions of the intrinsically disordered nuclear chromatin protein NUPR1. *Biomacromolecules* **2019**, *20*, 2567–2576. [CrossRef]
58. Yadahalli, S.; Neira, J.L.; Johnson, C.M.; Tan, Y.S.; Rowling, P.J.E.; Chattopadhyay, A.; Verma, C.; Itzhaki, L.S. Kinetic and thermodynamic effects of phosphorylation on p53 binding to MDM2. *Sci. Rep.* **2019**, *9*, 693. [CrossRef]
59. Bah, A.; Vernon, R.M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.W.; Sonenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2014**, *519*, 106–109. [CrossRef]
60. Gandhi, N.S.; Landrieu, I.; Byrne, C.; Kukić, P.; Amniai, L.; Cantrelle, F.-X.; Wieruszeski, J.-M.; Mancera, R.L.; Jacquot, Y.; Lippens, G. A Phosphorylation-induced turn defines the Alzheimer's disease AT8 antibody epitope on the Tau protein. *Angew. Chem. Int. Ed.* **2015**, *54*, 6819–6823. [CrossRef]
61. Beck, D.A.C.; Alonso, D.O.V.; Inoyama, D.; Daggett, V. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12259–12264. [CrossRef]
62. Muñoz, V.; Serrano, L. Intrinsic secondary structure propensities of the amino acids, using statistical φ-ψ matrices: Comparison with experimental scales. *Proteins: Struct. Funct. Bioinform.* **1994**, *20*, 301–311. [CrossRef]
63. Gibbs, E.B.; Lu, F.; Portz, B.; Fisher, M.J.; Medellin, B.P.; Laremore, T.N.; Zhang, Y.S.; Gimour, D.S.; Showalter, S.A. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* **2017**, *8*, 15233. [CrossRef] [PubMed]
64. Xiang, S.; Gapsys, V.; Kim, H.-Y.; Bessonov, S.; Hsiao, H.-H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; et al. Phosphorylation drives a dynamic switch in Serine/Arginine-rich proteins. *Structure* **2013**, *21*, 2162–2174. [CrossRef] [PubMed]
65. Turner, A.L.; Watson, M.; Wilkins, O.G.; Cato, L.; Travers, A.; Thomas, J.O.; Stott, K. Highly disordered histone H1−DNA model complexes and their condensates. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11964–11969. [CrossRef] [PubMed]

66. Banavali, N.K.; Roux, B. Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins: Struct. Funct. Bioinform.* **2007**, *67*, 1096–1112. [CrossRef]
67. Espinoza-Fonseca, L.M.; Kast, D.; Thomas, D.D. Molecular dynamics simulations reveal a disorder-to-order transition on phosphorylation of smooth muscle myosin. *Biophys. J.* **2007**, *93*, 2083–2090. [CrossRef]
68. Hendus-Altenburger, R.; Lambrughi, M.; Terkelsen, T.; Pedersen, S.F.; Papaleo, E.; Lindorff-Larsen, K.; Kragelund, B.B. A phosphorylation-motif for tuneable helix stabilisation in intrinsically disordered proteins–Lessons from the sodium proton exchanger 1 (NHE1). *Cell. Signal.* **2017**, *37*, 40–51. [CrossRef]
69. Chu, I.M.; Hengst, L.; Slingerland, J.M. The Cdk inhibitor p27 in human cancer: Prognostic potential and relevance to anticancer therapy. *Nat. Rev. Cancer* **2008**, *8*, 253–267. [CrossRef]
70. He, Y.; Chen, Y.; Mooney, S.M.; Rajagopalan, K.; Bhargava, A.; Sacho, E.; Weninger, K.; Bryan, P.N.; Kulkarni, P.; Orban, J. Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen*. *J. Boil. Chem.* **2015**, *290*, 25090–25102. [CrossRef]
71. Zimmermann, M.T.; Williams, M.M.; Klee, E.W.; Lomberk, G.L.; Urrutia, R.A. Modeling post–translational modifications and cancer–associated mutations that impact the heterochromatin protein 1α–importin α heterodimers. *Proteins: Struct. Funct. Bioinform.* **2019**, *87*, 904–916. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

MDPI

Article

The Paralogue of the Intrinsically Disordered Nuclear Protein 1 Has a Nuclear Localization Sequence that Binds to Human Importin α3

José L. Neira [1,2,*], Bruno Rizzuti [3], Ana Jiménez-Alesanco [2], Olga Abián [2,4,5,6,7], Adrián Velázquez-Campoy [2,4,5,6,8] and Juan L. Iovanna [9,*]

[1] IDIBE, Universidad Miguel Hernández, 03202 Elche (Alicante), Spain
[2] Instituto de Biocomputación y Física de Sistemas Complejos, Joint Units IQFR-CSIC-BIFI, and GBsC-CSIC-BIFI, Universidad de Zaragoza, 50009 Zaragoza, Spain; ajimenez@bifi.es (A.J.-A.); oabifra@unizar.es (O.A.); adrianvc@unizar.es (A.V.-C.)
[3] CNR-NANOTEC, Licryl-UOS Cosenza and CEMIF.Cal, Department of Physics, University of Calabria, Via P. Bucci, Cubo 31 C, Arcavacata di Rende, 87036 Cosenza, Italy; bruno.rizzuti@cnr.it
[4] Instituto de Investigación Sanitaria Aragón (IIS Aragón), 50009 Zaragoza, Spain
[5] Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd), 28029 Madrid, Spain
[6] Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, 50009 Zaragoza, Spain
[7] Instituto Aragonés de Ciencias de la Salud (IACS), 50009 Zaragoza, Spain
[8] Fundacion ARAID, Gobierno de Aragon, 50009 Zaragoza, Spain
[9] Centre de Recherche en Cancérologie de Marseille (CRCM), INSERM U1068, CNRS UMR 7258, Aix-Marseille Université and Institut Paoli-Calmettes, Parc Scientifique et Technologique de Luminy, 163 Avenue de Luminy, 13288 Marseille, France
* Correspondence: jlneira@umh.es (J.L.N.); juan.iovanna@inserm.fr (J.L.I.); Tel.: +34-966-65-8475 (J.L.N.); +33(0)491-82-8803 (J.L.I.)

Received: 6 September 2020; Accepted: 3 October 2020; Published: 8 October 2020

Abstract: Numerous carrier proteins intervene in protein transport from the cytoplasm to the nucleus in eukaryotic cells. One of those is importin α, with several human isoforms; among them, importin α3 (Impα3) features a particularly high flexibility. The protein NUPR1L is an intrinsically disordered protein (IDP), evolved as a paralogue of nuclear protein 1 (NUPR1), which is involved in chromatin remodeling and DNA repair. It is predicted that NUPR1L has a nuclear localization sequence (NLS) from residues Arg51 to Gln74, in order to allow for nuclear translocation. We studied in this work the ability of intact NUPR1L to bind Impα3 and its depleted species, ΔImpα3, withou the importin binding domain (IBB), using fluorescence, isothermal titration calorimetry (ITC), circular dichroism (CD), nuclear magnetic resonance (NMR), and molecular docking techniques. Furthermore, the binding of the peptide matching the isolated NLS region of NUPR1L (NLS-NUPR1L) was also studied using the same methods. Our results show that NUPR1L was bound to Imp α3 with a low micromolar affinity (~5 μM). Furthermore, a similar affinity value was observed for the binding of NLS-NUPR1L. These findings indicate that the NLS region, which was unfolded in isolation in solution, was essentially responsible for the binding of NUPR1L to both importin species. This result was also confirmed by our in silico modeling. The binding reaction of NLS-NUPR1L to ΔImpα3 showed a larger affinity (i.e., lower dissociation constant) compared with that of Impα3, confirming that the IBB could act as an auto-inhibition region of Impα3. Taken together, our findings pinpoint the theoretical predictions of the NLS region in NUPR1L and, more importantly, suggest that this IDP relies on an importin for its nuclear translocation.

Keywords: circular dichroism; fluorescence; importin; intrinsically disordered protein (IDP); isothermal titration calorimetry (ITC); molecular docking; nuclear magnetic resonance (NMR); paralogue; peptide

---

1. Introduction

NUPR1 (UniProtKB O60356) is an 82-residue-long (8 kDa), monomeric intrinsically disordered protein (IDP) with a large content of basic residues [1,2]. It does not have a stable secondary and tertiary structure, as also happens, at least partially, for other IDPs [3–5]. NUPR1 is involved in chromatin remodeling and transcription, and it is an important element in cell cycle regulation and cell stress response [6,7]. It is also implicated in apoptosis, forming a complex with another IDP, prothymosin α [8,9], as well as being involved in DNA binding and repair [10,11], and in the interaction with Polycomb group proteins [12]. Expression of the *NUPR1* gene is down-regulated by the presence of NUPR1L, a 97-residue-long paralogue of NUPR1; in turn, the expression of NUPR1L is p53-regulated [13]. We have recently shown that NUPR1L is also an IDP, but it has a higher tendency to self-associate than NUPR1 [14], and it shows regions with conformations including turn- or helix-like structures.

The active transport of proteins from the cytoplasm to the nucleus occurs through several transport receptors known as importins (or karyopherins), co-operating with other proteins such as GTPase Ran and nucleoporins [15–17]. The classical nuclear import pathway is initiated by recognition of a typical amino acid sequence (NLS, nuclear location sequence) in the cargo by an importin α [18]. The complex cargo-importin α binds to importin β, through the importin β-binding domain (IBB), and the ternary complex moves through the nuclear pore complex (NPC). The complex within the nucleus is dissociated by the action of GTPase Ran interacting with importin β, and both importins α and β are recycled back to the cytoplasm [18]. There are seven isoforms of importin α in humans, which have a role in cell differentiation, gene regulation, and cancer development [19,20]. We have chosen Impα3 as a target for NUPR1 because of its larger flexibility in comparison with other importins, as concluded from X-ray data, which confer it a greater ability to interact with different cargos. In addition, from a practical point of view, Impα3 can be also easily expressed and purified for in vitro structural studies. Interestingly, it has also been shown to be crucial in pain pathways [21].

Importin α is formed by two domains: (i) an N-terminal 60-residue-long IBB domain and (ii) a C-terminal NLS-binding motif formed by ten armadillo (ARM) repeat units [15,17,19,20]. The interaction with the cargo occurs in a concave site of the elongated structure, involving ARM motifs 2 to 4 (major site) or 6 to 8 (minor site) for the shortest monopartite NLSs, or both sets of ARM motifs for the largest bipartite NLS regions. If Importin β is not present, the IBB domain, mimicking an NLS region, occupies the same ARM motifs involved in NLS recognition, and then it has an intramolecular auto-inhibitory role [22].

We have previously shown that NUPR1 binds to human importin α3 (Impα3), also called KPNA4 [23]. NUPR1 has an NLS region involving residues in the 60–70 s along the protein sequence, as has been shown by molecular cell biology studies [24]. In this work, we studied the interaction of Impα3, and that of its truncated species without the IBB domain (ΔImpα3), with NUPR1L and with its predicted NLS region, NLS-NUPR1L (comprising residues Arg51 to Gln74), using several biophysical techniques, namely, fluorescence, circular dichroism (CD), isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) and molecular docking. Our results show that, as occurs with the parent NUPR1 [23], the intact NUPR1L was capable of interacting with both importin species with affinity in the low micromolar range (~5 μM). The NLS-NUPR1L was disordered in solution when it was in isolation, but it was bound to both importin species with similar affinity as the intact NUPR1L, suggesting that this protein region contains all the key residues determining the binding.

In all cases, the affinity for ΔImpα3 was larger than for intact importin, indicating that the IBB has an auto-inhibitory effect for binding any cargo.

2. Results

2.1. Intact NUPR1L is Associated with Both Impα3 and ΔImpα3

We first determined whether intact NUPR1L could bind to Impα3 and ΔImpα3, keeping in mind that this IDP has a high tendency to aggregate [14]. We first mapped, by fluorescence and CD, whether there was binding between NUPR1L and each importin species, by comparing changes in the fluorescence and far-UV CD spectra of the complex with those obtained from the sum of the corresponding spectra of each molecule. Although the far-UV CD spectra are dominated by the presence of importins, because of their larger size and higher number of peptide bonds (when compared with that of isolated NUPR1L), the spectra can provide valuable information. Figure 1 shows the spectra obtained with ΔImpα3 (the results for Impα3 are provided in the Supplementary Material, Figure S1). The results of CD and fluorescence for ΔImpα3 indicate the following: (i) there were changes in the environment around tryptophan residues of at least one of the two proteins (i.e., NUPR1L or ΔImpα3) upon binding (fluorescence spectra, Figure 1A), and (ii) there were changes in the secondary structure of at least one of the proteins upon binding (CD spectra; Figure 1B). As NUPR1L has no well-defined structure [14], and ΔImpα3 is a large protein with a rigid, well-formed helical fold [25], we suggest that the changes in CD spectra were due to the acquisition of structure by NUPR1L. We did not attempt to deconvolute the spectrum of the complex because of the presence of two polypeptide chains, and the fact that we do not know the exact conformation of NUPR1L. We further carried out thermal denaturations followed by CD; as can be observed (Figure 1C), the apparent thermal midpoint of the unfolding of ΔImpα3 changed from 317 K to 323 K, indicating the presence of binding (leading to a stabilization of the folded state of ΔImpα3). We did not follow the binding by changes in the thermal denaturation midpoint, as monitored by fluorescence, as the sigmoidal curves of both isolated importin species obtained with this technique are not as clearly defined as those from CD [25].

Next, we tried to use ITC to determine quantitatively the binding parameters between NUPR1L and the two importins, as ITC is the gold-standard in measuring thermodynamic parameters of any binding reaction. However, a large peak observed in the thermograms upon dilution of NUPR1L precluded any measurement of the binding to the importin species contained in the cell. Then, we tried to measure the binding of NUPR1L to both importins using fluorescence (Figure 2), keeping its concentration in the cuvette constant. NUPR1L bound to both importins, with similar apparent dissociation constants, in the low micromolar range: (4.0 ± 0.7) μM for Impα3 (Figure 2A), and (5 ± 1) μM for ΔImpα3 (Figure 2B). We observed that the data for ΔImpα3 were more scattered; although we do not have a clear explanation for this finding, it might be due to the lower solubility of ΔImpα3 [25].

2.2. Isolated NLS-NUPR1L Was Bound to Impα3 and ΔImpα3

As intact NUPR1L associated to both importins, we wondered whether (i) the isolated predicted NLS region was capable of binding to them as well, and (ii) the affinity was the same as that of the intact protein. To that end, we first determined the conformational preferences of the isolated NLS-NUPR1L region by several spectroscopic methods.

2.2.1. Isolated NLS-NUPR1L Was Monomeric and Disordered in Solution

The fluorescence spectrum of the peptide had a maximum at 353 nm (Figure S2), close to the wavelength where the maximum of fluorescence for a solvent-exposed tryptophan is expected [26]; therefore, we could conclude that the sole tryptophan present in NLS-NUPR1L (Trp62) was exposed to the solvent. The CD spectrum of isolated NLS-NUPR1L did show an intense minimum at ~203 nm (Figure 3A), indicating that the peptide acquired a random-coil conformation. This was further confirmed by 1D-^{1}H-NMR spectra (Figure 3B), which showed a clustering of the signals of all the amide protons between 8.0 and 8.5 ppm, and the methyl protons were observed between 0.8 and 1.0 ppm, which is a feature typical of disordered polypeptide chains [27].

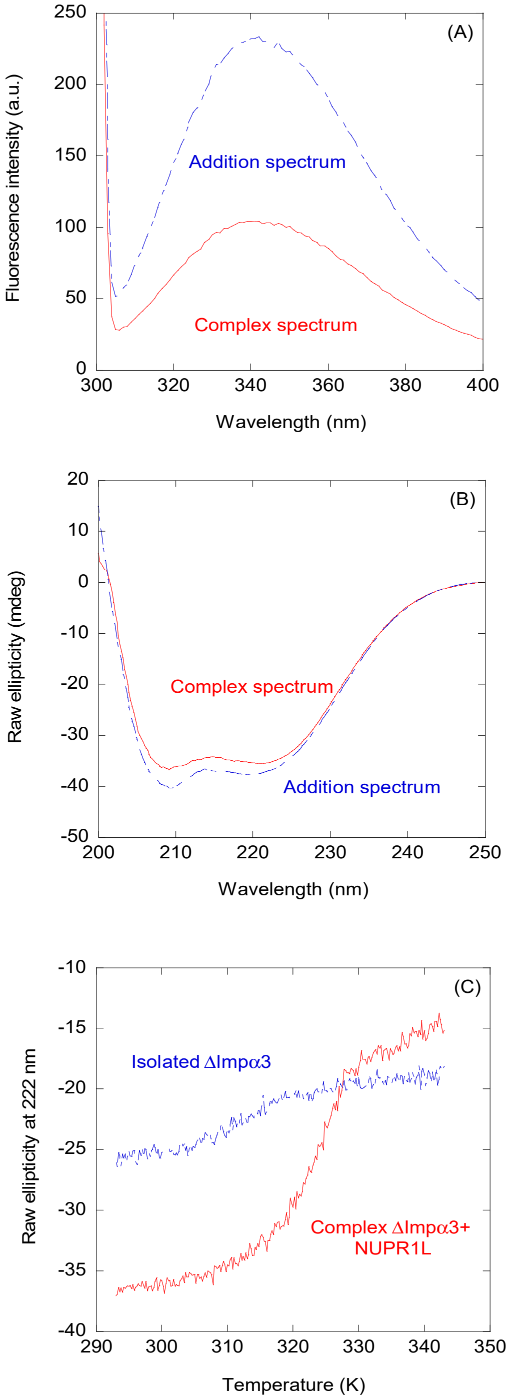


Figure 1. Binding of intact nuclear protein 1 NUPR1L to ΔImpα3 monitored by spectroscopic techniques: (**A**) Fluorescence spectrum obtained by excitation at 295 nm of the complex between ΔImpα3 and intact NUPR1L, and addition spectrum obtained by the sum of the spectra of both isolated macromolecules. (**B**) Far-UV circular dichroism (CD) spectrum of the complex between ΔImpα3 and NUPR1L and the addition spectrum obtained by the sum of the spectra of both isolated macromolecules. (**C**) Thermal denaturations of ΔImpα3 in the presence and absence of NUPR1L followed by the changes in ellipticity at 222 nm. All experiments were carried out in phosphate buffer (50 mM, pH 7.0).

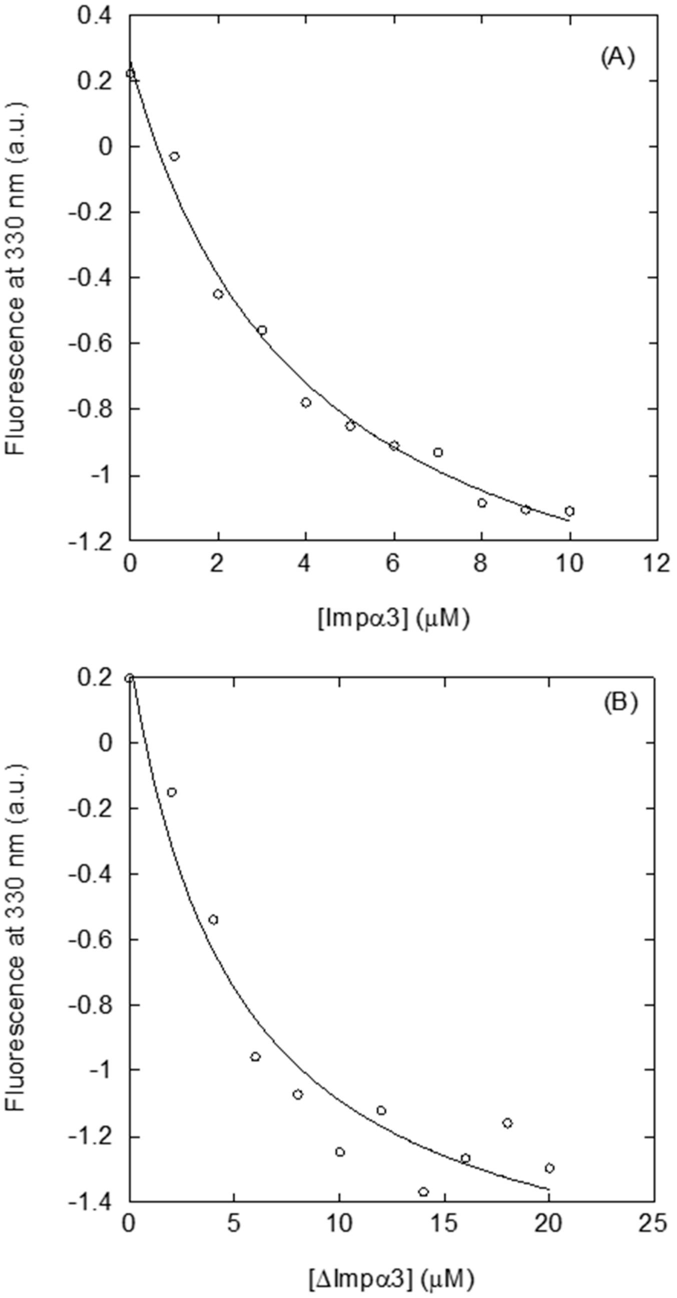


Figure 2. Interaction of intact NUPR1L with both importin species as measured by fluorescence: (**A**) Titration curve monitoring the changes of Impα3 fluorescence at 330 nm in the presence of NUPR1L, after excitation at 280 nm. (**B**) Titration curve monitoring the changes of ΔImpα3 fluorescence at 330 nm in the presence of NUPR1L, after excitation at 280 nm. All experiments were carried out in phosphate buffer (50 mM, pH 7.0).

The peptide was monomeric, as concluded from the value of D measured by the DOSY (diffusion ordered spectroscopy) and the estimated R_h obtained from comparison with that of dioxane: (1.74 ± 0.05) $\times 10^{-6}$ cm^2 s^{-1} and (12 ± 2) Å, respectively. This value of R_h was similar to that obtained theoretically for a random-coil polypeptide [28]: 14 ± 3 Å.

To further confirm the disordered nature of NLS-NUPR1, we also carried out homonuclear 2D-^{1}H-NMR experiments (Table S1). We observed NOEs between the H_α protons of Trp62 and the H_δ of Pro63, but we also observed other signals involving residues around Trp62; in fact, two signals were observed for the indole proton of Trp62 (Figure 1B) (Table S1). These results indicate the presence of the cis-trans equilibrium between the two conformations of Pro63, probably favored by the bulkiness of the side-chain of Trp62. The peptide was mainly disordered in solution, as suggested by two lines of evidence (further pinpointing the results from fluorescence (Figure S2), far-UV CD (Figure 3A), and 1D-^{1}H-NMR spectra (Figure 3B). First, the sequence-corrected conformational shifts ($\Delta\delta$) of H_α protons [27,29,30] were within the commonly accepted range for random-coil peptides ($\Delta\delta \leq 0.1$ ppm) (Table S1). Second, no long- or medium-range NOEs were detected, but only sequential ones (Figure 3C).

To sum up, all the experimental techniques concurred to indicate that the isolated NLS-NUPR1L was disordered in aqueous solution.

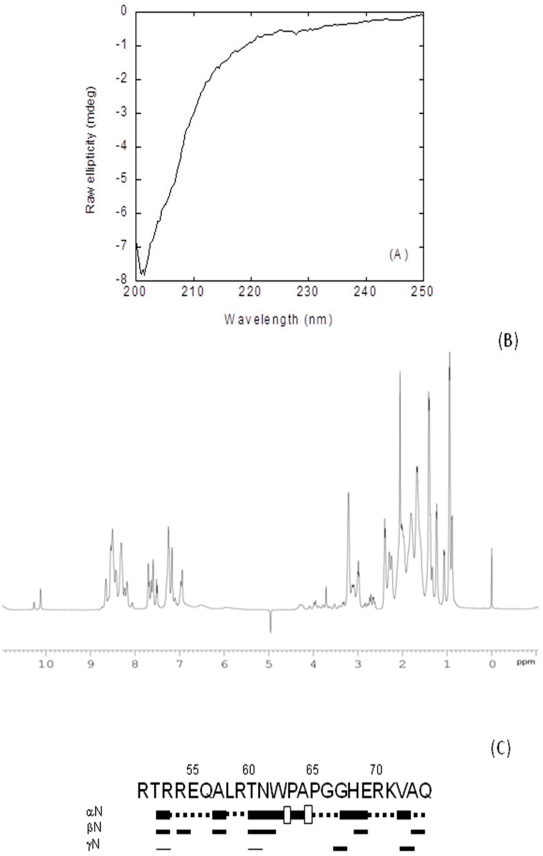


Figure 3. Conformational features of isolated nuclear localization sequence (NLS)-NUPR1L in solution: (**A**) Far-UV CD spectrum of NLS-NUPR1L at 298 K in phosphate buffer (50 mM, pH 7.0). (**B**) 1D-^{1}H-nuclear magnetic resonance (NMR) spectrum of isolated NLS-NUPR1L at 283 K and pH 7.2 (50 mM, Tris buffer). (**C**) NOE (Nuclear Overhuaser effect) diagram of isolated NLS-NUPR1L at 283 K: NOEs are classified into strong, medium, or weak, as represented by the height of the bar underneath the sequence; signal intensity was judged by visual inspection from the NOESY (Nuclear Overhuaser effect spectrosocopy) experiments. The corresponding H_α NOEs with the H_δ of the following proline residue are indicated by an open bar in the row corresponding to the sequential αN contacts. The dotted lines indicate NOE contacts that could not be unambiguously assigned owing to signal overlap. The numbering of residues corresponds to that of the sequence of intact NUPR1L. The symbols αN, βN, γN, and NN correspond to the sequential contacts (that is, for instance, the αN corresponds to the αN (i,i + 1) contacts).

2.2.2. Isolated NLS-NUPR1L Associated with Both Importins

In the following step, we measured the affinity of NLS-NUPR1L for both importins. We followed the same procedure as with intact NUPR1L, that is, first we tried to detect changes using fluorescence;

CD; and, in this case, T_2-relaxation measurements; next, we measured quantitatively the affinity using ITC and fluorescence.

Fluorescence and CD experiments showed that there were changes in the spectra upon addition of NLS-NUPR1L to each of the importin species (Figures S3 and S4), although again, the CD spectra are dominated by signal of the importins, and the effect is more marked in this case as a result of the smaller size of NLS-NUPR1L compared with the whole NUPR1L. The changes for ΔImpα3 (Figure S4) were similar (either in the steady-state spectra for both techniques or in the thermal denaturations followed by CD) to those observed for the intact NUPR1L with ΔImpα3 (Figure 1). Experiments aimed to detect binding using relaxation NMR measurements were only carried out with Impα3, because of its larger solubility [25]. It is important to note that the T_2 is greater for small molecules and shorter in larger molecules (or complexes) owing to a larger number of dipole–dipole interactions [31]. The T_2 of the most up-field shifted indole signal (that is, the one with the larger intensity) was measured in the isolated peptide, and was 62.2 ms; conversely, the T_2 in the presence of Impα3 was 36.7 ms, in agreement with what we should expect upon complex formation [31].

Then, we proceeded to determine the binding between the same molecules using fluorescence and ITC. The fluorescence results (Table 1, Figure 4A,B) yielded values similar to those measured for the intact NUPR1L (Section 2.1), but ITC yielded a dissociation constant larger for Impα3 (12 μM), and a value similar to that measured from fluorescence for ΔImpα3 (5 μM) (Table 1, Figure 4C). Similar discrepancies in the measured affinity constants among different techniques have been observed when measuring interactions in other proteins [32–35]. The reason behind such discrepancy is related to the particular features of each technique. Steady-state techniques, where the physical observable is the equilibrium state after long incubation times that allow an optimal accommodation of the interacting molecules (such as fluorescence titration), may provide higher affinities than transient-event techniques, where the observable quantity mainly reflects the first encounter between the interacting molecules (such as ITC), thus kinetically slow readjusting conformational events may be overlooked. Because the stoichiometry of binding is already accounted for in the binding model and both importins slightly differ in the parameter n, that difference in the parameter n for both importins could be due to the lower solubility of ΔImpα3 [25], resulting in a lower fraction of active or binding-competent protein.

Table 1. Thermodynamic parameters at 298 K in the binding reaction of nuclear localization sequence (NLS)-NUPR1L to the two importin species. NUPR1, nuclear protein 1; ITC, isothermal titration calorimetry.

	Fluorescence	**ITC**		
Importin Species	**K_d (μM)**	**K_d (μM)**	**ΔH (kcal mol^{-1})**	***n***
Impα3	3 ± 1	12 ± 2	−3.1 ± 0.5	1.04 ± 0.05
ΔImpα3	5 ± 2	5.5 ± 0.9	−2.4 ± 0.5	0.75 ± 0.06

2.2.3. Binding Regions in the Docking of NLS-NUPR1L to Importins

Molecular docking was used to predict the binding location of NLS-NUPR1L on the surface of Impα3, and to clarify the structural basis of their interactions. Because of the relatively high number of degrees of freedom of the 24-residue-long peptide used in our experiments and its large structural flexibility, our in silico research was carried out considering nine 8-residue-long fragments of this peptide, each possessing a number of rotatable bonds (ranging from 23 to 38, depending on the fragment) close to the limit considered reliable to be computationally tractable by the docking engine [36]. Any possible bias in the simulation was avoided by performing a blind docking on the whole protein volume and using a very high exhaustiveness in the search.

Figure 5 summarizes the binding affinity for ΔImpα3 of the peptide fragments, as obtained in the docking experiments. The energy value for each of the 8-residue-long fragment is reported in correspondence to its two central amino acids. We also verified that the computational depth in the docking search was reasonable to obtain a statistical convergence of the binding score values obtained (inset of Figure 5), indicating that the conformational space for the 8-residue-long fragments could be considered exhaustively sampled. The most favorable binding score was observed for the fragment with sequence PAPGGHER, which is one of the regions with the highest conformational flexibility in the parent peptide sequence due to the presence of two couples of disorder-prone Pro and Gly residues. As discussed above, this region is probably responsible for hampering the free rotation of the indole moiety of Trp62, and then of the presence of both indole signals (Figure 3B), and thus two distinct conformers. The binding energy of the fragment was −7.9 kcal/mol, indicating an affinity in the low micromolar range. The predicted core region of the NLS of NUPR1L essentially maps in correspondence with that of NUPR1 [24], although the former is shifted a few residues towards the N-terminal region of the main chain compared with the latter, when the two protein sequences are aligned. More generally, all the fragments that were part of the 14-residue-long sequence RTNWPAPGGHERKV showed energies ≤ −7.5 kcal/mol, suggesting that this whole region may contribute to the binding of NUPR1L to Impα3.

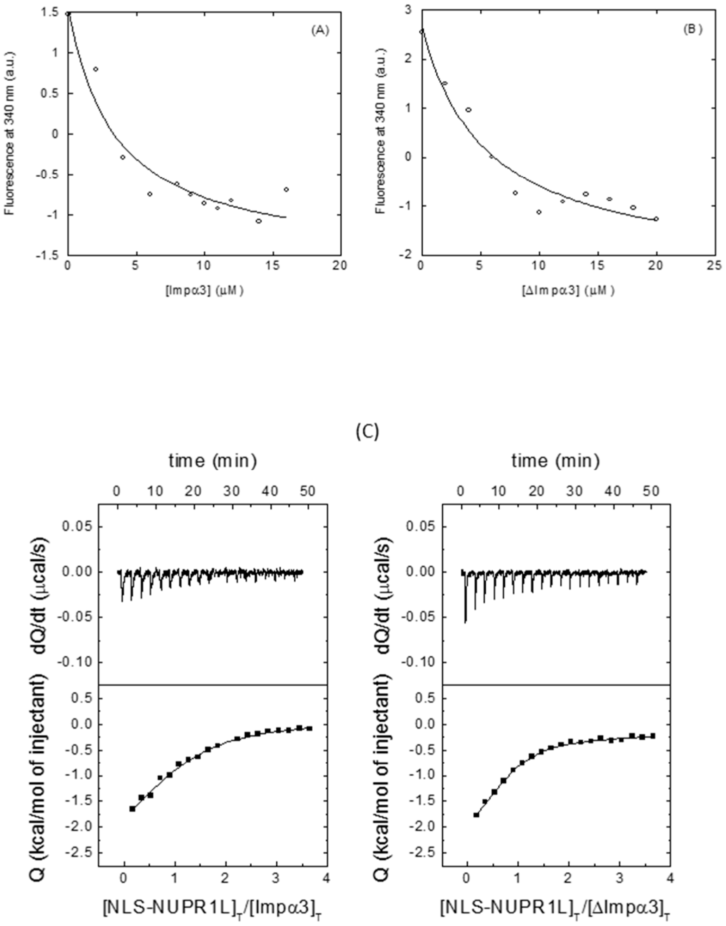


Figure 4. Binding of NLS-NUNPR1L to both importin species: (**A**) Titration curve monitoring the changes of Impα3 fluorescence at 340 nm in the presence of NLS-NUPR1L, after excitation at 280 nm. (**B**) Titration curve monitoring the changes of ΔImpα3 fluorescence at 340 nm in the presence of NLS-NUPR1L, after excitation at 280 nm. All experiments were carried out in phosphate buffer (50 mM, pH 7.0). (**C**) Calorimetric binding isotherms (ligand normalized heat effect per injection as a function of the ligand/protein molar ratio) for the interaction of NLS-NUPR1L with Impα3 (left) and ΔImpα3 (right) are shown, with the thermogram (raw thermal power data as a function of time) at the top of each panel. Binding parameters were estimated by non-linear least squares regression data analysis of the interaction isotherms applying a single ligand binding site model implemented in Origin 7.0. (OriginLab, Northampton, MA, USA).

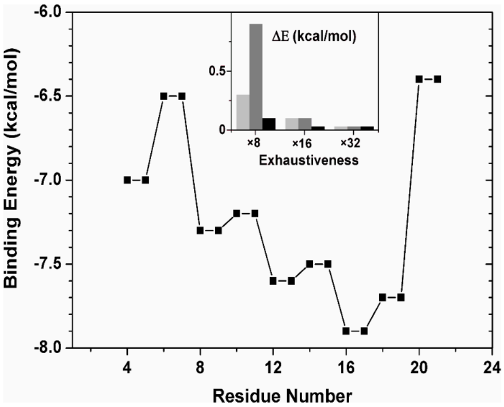


Figure 5. Binding energy of 8-residue-long fragments of NLS-NUPR1L peptide to importin: Affinity of each fragment is shown in correspondence of the two residues at the centre of each 8-residue sequence. (Inset) Difference in the binding energy between the docking pose with the highest affinity found at increasing exhaustiveness in the search versus the best pose found at any exhaustiveness value, for three 8-residue-long fragments that together span the whole sequence of the twenty-residue-long NLS-NUPR1L: (light grey) fragment RTRREQAL, (dark grey) RTNWPAPG, and (black) GHERKVAQ.

We also performed a structural analysis of the interaction between the NUPR1L sequence fragments and ΔImpα3. Figure 6A shows the best five binding modes obtained for the polypeptide fragment PAPGGHER, which includes the most favorable docking pose and other four poses with binding scores within 0.3 kcal/mol. All the fragment conformations were found to cluster in correspondence with the ARM repeats 2–4 of Impα3 (which are also present in ΔImpα3), which corresponds to the major binding site for the NLS of cargo proteins (Section 1). Moreover, as detailed in Figure 6B, the most favorable conformation of the fragment PAPGGHER was found to overlap with the NLS of the EBNA-LP protein of the Epstein–Barr virus [37], that is, the crystallographic ligand complexed with Impα3 in the protein structure used for the docking experiments. Key residues in the interaction with the NLS of NUPR1L were the tryptophans in the major binding site of Impα3 (labeled in Figure 6B), which are known to be essential in maintaining the binding with the EBNA-LP protein and in other importin–cargo complexes. A number of other amino acids of Impα3 also participated in the binding, including residue Asp192, which forms a salt bridge with the arginine residue in the peptide, Arg70, close to the C terminus of the fragment of NLS-NUPR1L. However, we cannot exclude that other electrostatic interactions occur in the intact NLS region with nearby arginines such as Arg59, Arg54, or Arg53.

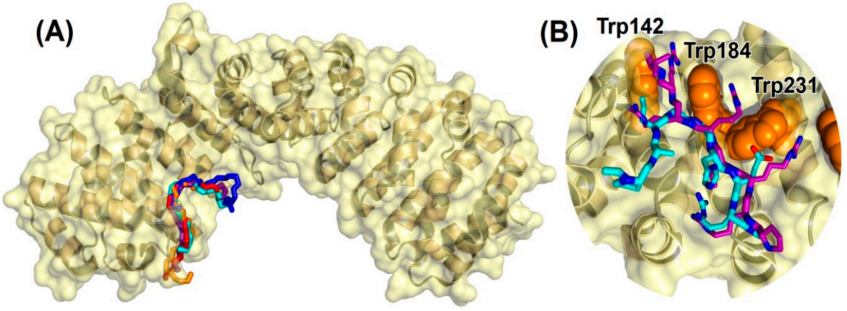


Figure 6. Predicted docking poses for NUPR1L peptide on importin. (**A**) Bound conformation of the capped fragment PAPGGHER: backbone (–N–C$^{\alpha}$–C– atoms) representation of the best five docking poses on ΔImpα3. (**B**) Most favorable binding pose of the same fragment (cyan), compared with the crystallographic conformation [37] of the NLS of the Epstein–Barr virus EBNA-LP protein (purple). For clarity, H atoms and backbone O atoms are omitted. The tryptophan residues (orange) in the major NLS-binding site of importin are labeled. All images were created with PyMol [38].

3. Discussion

3.1. Identification of the NLS Region of NUPR1L

The first result of our work is that NUPR1L contains an NLS region, which is responsible for its binding to Impα3. Furthermore, the isolated region binds to both importin species with nearly the same affinity as to the whole IDP, indicating that all the key amino acids responsible for binding are mainly contained in such a polypeptide patch. Therefore, importins are capable of binding to both paralogues, NUPR1L and NUPR1, as we had already demonstrated the binding to the latter protein [23]. These findings are at variance with recent results obtained by fluorescence, where other IDPs were suggested to translocate into the nucleus without the need for the nucleus-cytoplasm transport machinery [39].

The docking simulation makes clear a number of points about the interaction between NUPR1L and Impα3. We identified a core region of the NLS of the NUPR1L, which is located in the same regions of the NLS already predicted and validated for the paralogue, NUPR1 [23,24]. Such a region of NUPR1L is a hot spot that binds to the major NLS-binding site of Impα3, located in the ARM repeats 2–4. The bound conformation of NLS-NUPR1L overlaps with the one obtained in crystallography for the NLS of the Epstein–Barr virus EBNA-LP protein [37] (Figure 6B). Hydrophobic interactions with the tryptophan residues in the binding site of Impα3 were crucial for the binding, suggesting a common cargo-binding mechanism shared by well-folded proteins and IDPs. In addition, these in silico results support our fluorescence findings, as it was possible to measure the binding between each importin and either the intact protein or the isolated peptide, by following the changes in fluorescence of both at 280 or 295 nm (Figures 1 and 4A,B), indicating the tryptophans were involved in the reaction. Electrostatic interactions with the charged residue Asp192 of Impα3 provide a further anchor for Arg70 in the NLS of NUPR1L, contributing to securing its bound position the outmost C-terminal region of the NUPR1L peptide used in our experiments.

We also demonstrated that isolated NLS-NUPR1L did not have any propensity to acquire helix- or turn-like conformations; this result is important as we have previously shown that NUPR1L has a tendency to form locally folded regions around Trp62 [14], and with the present findings, we can conclude that that folded conformation around this region was not helical (Table S1). In this aspect, NLS-NUPR1L behaves not differently from any other NLS region of a well-folded protein [15,20,22,40]; that is, it is disordered both in isolation and when participating in forming the complex with importins (in our studies, the latter conclusion was obtained from our docking simulations). Finally, it is important to pinpoint that Trp62 is also involved in the binding of NUPR1L to prothymosin α [14]; therefore, it seems that this residue can be classified as a hot spot of NUPR1L in the association with other molecular partners.

3.2. The Inhibitory Effect of the IBB in Impα3

We can conclude (Table 1 and Section 2.2.2) that the removal of IBB from Impα3 promotes a more favorable binding of the NLS-NUPR1 to the ARM 2–3 units of importins: the dissociation constants were 5.5 μM (for ΔImpα3) versus 12 μM (for Impα3). Unfortunately, we cannot draw any defined conclusion for the intact NUPR1L, as we could not measure the binding parameters by ITC (Section 2.1). Nevertheless, the result obtained is in agreement with previous findings of other NLS regions of well-folded proteins [40] or with those of the intact NUPR1 protein (1.4 μM for Impα3) [23] or peptides comprising the NLS region of NUPR1 [41]. The presence of the IBB (which contains a large quantity of lysine amino acids) always exerts an auto-inhibitory effect, and the domain hampers the anchoring of NLS-NUPR1L into the major NLS-binding region of Impα3. The modulation of the complex formation between importins and their cargos (belonging to otherwise well-folded proteins) has been attributed to the IBB [18]; interestingly enough, this region is involved even in the formation of a homodimeric species between importins [42], conferring to this protein a reduced ability to bind cargos.

3.3. Binding to Impα3 of NUPR1L

As the isolated NLS regions of both NUPR1 and NUPR1L contain the key residues to attain binding to Impα3, and with the peptides, we could measure the binding by ITC, we shall focus our attention on the comparison between the affinities of the two paralogues for those measurements. Comparison of the values of Table 1 for NLS-NUPR1L with those of the NLS region of NUPR1 (1.7 μM for Impα3 and 0.95 μM for ΔImpα3 [41]) indicate that the binding is stronger in the case of NUPR1. Therefore, although both paralogues bind to the same molecules ([14] and this work), their affinity for the different partners is dissimilar. This could provide a mechanism to explain the regulation between these proteins, not only at a DNA level, but also at a post-translational stage.

In the case of the intact proteins, although we do not have the whole set of values of K_d obtained with the same technique (for NUPR1L, the dissociation constants were obtained by fluorescence (Section 2.1), whereas for NUPR1, they were obtained by ITC [23,41]), it is important to consider that NUPR1L in solution is an oligomer and, therefore, the self-association equilibrium will affect the apparent values of the dissociation constants determined from the experiments, which could vary depending on the self-association state of the protein.

4. Materials and Methods

4.1. Materials

Ampicillin and isopropyl-β-D-1-tiogalactopyranoside were from Apollo Scientific (Stockport, UK). Imidazole, kanamycin, TSP ((trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid), Trizma base, and His-Select HF nickel resin were from Sigma-Aldrich (Madrid, Spain). Triton X-100 and protein marker (PAGEmark Tricolor) were from VWR (Barcelona, Spain). Amicon centrifugal devices with a cut-off molecular weight of 30 or 50 kDa were from Millipore (Barcelona, Spain). The rest of the materials were of analytical grade. Water was deionized and purified on a Millipore system.

4.2. Protein Expression and Purification

Expression and purification of codon-optimized, His-tagged ΔImpα3 (residues 64-521) were carried out using BL21 (DE3) cells [25,40]. The DNA of the codon-optimized, intact Impα3 was synthesized by NZYtech (Lisbon, Portugal) and cloned into the pHTP1 vector (kanamycin resistance), and with a His-tag at the protein N terminus. Expression and purification of Impα3 were carried out as those for ΔImpα3 in the same *E. coli* strain cells. The protein concentration of both species was determined from their six tyrosines and six tryptophans [43]. NUPR1L was expressed and purified as described [14], and its concentration was determined from its single tryptophan and its five tyrosines [43].

4.3. Prediction and Synthesis of NLS-NUPR1L

The NLS-NUPR1L peptide was synthesized by NZYtech with a purity of 95%. The NLS region of NUPR1L was predicted using the whole sequence of NUPR1L in the web server http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi [44,45]. The predicted region with a larger score comprised residues Gly46 to Gln74. The peptide was designed to maximize solubility, comprising residues Arg51 to Gln74, with acetylation and amidation at the N and C termini, respectively, to avoid fraying effects.

4.4. Fluorescence

4.4.1. Steady-State Fluorescence

Fluorescence spectra were collected on a Cary Varian spectrofluorometer (Agilent, Santa Clara, CA, USA), interfaced with a Peltier unit. All experiments were carried out at 298 K. Following the standard protocols used in our laboratories, the samples were prepared the day before and left overnight at

278 K; before experiments, samples were left for 1 h at 298 K. A 1 cm pathlength quartz cell (Hellma, Kruibeke, Belgium) was used. The concentration of NLS-NUPR1L was 10 μM and those of both importins were 4 μM. Samples containing the isolated peptide, the isolated importin species, and a mixture of both (at those indicated concentrations) were prepared. Experiments were acquired in 50 mM phosphate buffer (pH 7.0). For the experiments with intact NUPR1L, a concentration of 15 μM (in protomer units) was used and that of each importin was 5 μM.

Protein samples were excited either at 280 or 295 nm. The other experimental parameters and the buffers used have been described elsewhere [46]. Appropriate blank corrections were made in all spectra.

4.4.2. Binding Experiments

For the titration between either Impα3 or ΔImpα3 with NUPR1L, increasing amounts of both importins, in the range 0–10 μM, were added to a solution with a fixed concentration of the intact IDP (8 μM). To maintain consistency, the same experimental set-up was used for titration of NLS-NUPR1L with both importins, although the peptide did not have any tendency to aggregate (Section 2.2.1); a fixed concentration of 8.5 μM of peptide was used in the titrations. Experiments were carried out in 50 mM buffer phosphate (pH 7.0) at 298 K. In all cases, the appropriate blank-corrections with the corresponding amounts of each importin species were subtracted. Spectra were corrected for inner-filter effects during fluorescence excitation [47]. Each titration (Impα3 with NUPR1L, Impα3 with NLS-NUPR1L, ΔImpα3 with NUPR1L, and ΔImpα3 with NLS-NUPR1L) was repeated at least three times, using new samples.

The samples were prepared the day before and left overnight at 278 K; before measurements, the samples were incubated for 1 h at 298 K. The dissociation constant of the corresponding complex, K_d, was calculated by fitting the binding isotherm obtained by plotting the observed fluorescence change as a function of importin concentration to the general binding model explicitly considering ligand depletion [48,49]:

$$F = F_0 + \frac{\Delta F_\mathrm{max}}{2[NUPR1L - polypep]_T}\left[\begin{array}{l}\left([NUPR1L - polypep]_T + [\mathrm{Imp}\alpha 3 - \mathrm{species}]_T + K_d\right) \\ -\left(\begin{array}{l}\left([NUPR1L - polypep]_T + [\mathrm{Imp}\alpha 3 - \mathrm{species}]_T + K_d\right)^2 \\ -4[NUPR1L - polypep]_T[\mathrm{Imp}\alpha 3 - \mathrm{species}]_T\end{array}\right)^{1/2}\end{array}\right] \quad (1)$$

where F is the measured fluorescence at any particular concentration of Impα3 or ΔImpα3 after subtraction of the blank with the same concentration of either Impα3 or ΔImpα3; ΔF_max is the largest change in the fluorescence of NUPR1L or NLS-NUPR1L when the whole amount of each polypeptide formed the complex compared with the fluorescence of each isolated chain; F_0 is the fluorescence intensity when no importin species was added; $[\mathrm{NUPR1L\text{-}polypep}]_T$ is the constant, total concentration of either NUPR1L or NLS-NUPR1L; and $[\mathrm{Imp}\alpha 3\text{-}\mathrm{species}]_T$ is that of either Impα3 or ΔImpα3, which was varied during the titration. Fitting to the above equation was carried out using KaleidaGraph version 3.5. (Synergy software, Reading, PA, USA).

4.5. CD

Far-UV CD spectra were collected on a Jasco J810 spectropolarimeter (Jasco, Tokyo, Japan) with a thermostated cell holder, and interfaced with a Peltier unit at 298 K. The instrument was periodically calibrated with (+)-10-camphorsulphonic acid. A cell with a path length of 0.1 cm was used (Hellma, Kruibeke, Belgium). All spectra were corrected by subtracting the corresponding baseline. The concentration of each polypeptide (importin species and either NLS-NUPR1L or intact NUPR1L) was the same as that used in the fluorescence experiments (Section 4.4).

4.5.1. Far-UV CD Spectra

Isothermal wavelength spectra of each isolated macromolecule and that of the complex were acquired with five scans at a scan speed of 50 nm/min, a response time of 2 s, and a band-width of 1 nm. Samples were prepared the day before and left overnight at 278 K to allow for equilibration. Before starting the experiments, samples were further left for 1 h at 298 K. Experiments were carried out at 298 K in 50 mM buffer phosphate (pH 7.0).

4.5.2. Thermal Denaturations

The experiments were performed at heating rates of 60 K/h and a response time of 8 s. Thermal scans were collected by following the changes in ellipticity at 222 nm typically from 298 to 343 K. The rest of the experimental set-up was the same as that reported in the steady-state experiments. No difference was observed between the scans aimed to test drifting in the signal of the spectropolarimeter. Thermal denaturations were not reversible for any of the polypeptides or their complexes, as shown by the following: (i) comparison of spectra before and after heating; and (ii) changes in the voltage of the instrument detector [50]. The apparent thermal denaturation midpoint was estimated from a two-state equilibrium equation, as previously described [46].

4.6. ITC

The experimental set-up and data processing of ITC experiments have been described previously [51]. Impα3 or ΔImpα3 (at 10–20 μM) was loaded into the cell of an Auto-iTC200 calorimeter (MicroCal, Malvern-Panalytical, Malvern, UK) and NLS-NUPR1L in the syringe (150–300 μM) in buffer Tris 50 mM, pH 8. The temperature for all experiments was 298 K. The experiments were analyzed applying a model considering a single ligand binding site (1:1 stoichiometry for the NLS-NUPR1L/Impα3 (or ΔImpα3) interaction) implemented in Origin 7.0 (OriginLab, Northampton, MA, USA).

4.7. NMR

The NMR experiments were acquired at 283 K on a Bruker Avance spectrometer (Bruker GmbH, Karlsruhe, Germany), equipped with a triple resonance probe and z-pulse field gradients. All experiments with NLS-NUPR1L were carried out at pH 7.2, 50 mM deuterated Tris buffer (not corrected for isotope effects). The spectra were calibrated with TSP ((trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid), by considering pH-dependent changes of its chemical-shifts [31]; probe temperature was calibrated with methanol [31].

4.7.1. D-^{1}H-NMR Spectrum

An amount of 128 scans was acquired with 16 K acquisition points for the homonuclear 1D-^{1}H-NMR spectrum, and using a peptide concentration of 1.0–1.2 mM. Water signal was suppressed using the WATERGATE sequence [52]. The spectrum was processed with Bruker TopSpin 2.1 (Bruker GmbH, Karlsruhe, Germany), after zero-filling and apodization with an exponential window.

4.7.2. Translational Diffusion NMR (DOSY)

The peptide concentration in DOSY experiment was 120 μM, and 128 scans were acquired, where the gradient strength was varied linearly. Translational self-diffusion measurements were performed with the pulsed-gradient spin-echo sequence in the presence of 100% D_2O. Experimental details have been described elsewhere [46]. The gradient strength was varied in sixteen linear steps between 2 and 95% of the total power of the gradient coil. The gradient strength was calibrated using the value of the translational diffusion coefficient, D, for the residual proton water signal in a sample containing 100% D_2O in a 5 mm tube [53]. The length of the gradient was 2.25 ms, the time between the two pulse gradients in the pulse sequence was 200 ms, and the recovery delay between the bipolar gradients was 100 µs. The methyl groups with signals between 1.0 and 0.80 ppm were used for peak integration (Section 2.2.1). Fitting of the exponential curves, obtained from experimental data as previously described [46], was carried out with KaleidaGraph version 3.5 (Synergy Software, Reading, PA, USA). A final concentration of 1% of dioxane, which was assumed to have a hydrodynamic radius R_h = 2.12 Å [53], was added to the peptide solution.

4.7.3. D-^{1}H-NMR Spectroscopy

Two-dimensional spectra were acquired in each dimension in the phase-sensitive mode using the time-proportional-phase incrementation technique (TPPI) and a spectral width of 7801.69 Hz [54]; the final concentration of the NLS-NUPR1L was the same as that used in the 1D experiments. Standard TOCSY (Total correlation spectroscopy) (with a mixing time of 80 ms) [55] and NOESY experiments (with a mixing time of 250 ms) [56] were performed by acquiring a data matrix size of 4096 × 512 points. The DIPSI (decoupling in the presence of scalar interactions) spin-lock sequence [57] was used in the TOCSY experiments with 1 s of relaxation time. Typically, 96 scans were acquired per increment in the first dimension, and the residual water signal was removed using the WATERGATE sequence [52]. NOESY spectra were collected typically with 96 scans per increment in the first dimension, with the residual water signal removed again by the WATERGATE sequence [52], and with 1 s of relaxation time. Data were zero-filled, resolution-enhanced with a square sine-bell window function optimized in each spectrum, baseline-corrected, and processed with the Bruker TopSpin 2.1 software (Bruker GmbH, Karlsruhe, Germany). The ^{1}H resonances were assigned by standard sequential assignment processes [27]. The chemical shift values of H_α protons in random-coil regions were obtained from tabulated data, corrected by neighbouring residue effects [27,29,30].

4.7.4. Measurements of T_2

Measurements of the T_2 (transverse relaxation time) provide a convenient method to determine the molecular mass of a macromolecule, as the correlation time, τ_c, is approximately equal to $1/(5 \times T_2)$ [58]. We measured the T_2 of one of the indole protons (Section 2.2.1) for NLS-NUPR1L (at 35 µM concentration) in isolation and in the presence of Impα3 (at a final concentration of 7 µM) with the 1-1 echo sequence [59]. The calculation of the T_2 was carried out as described [58].

4.8. Molecular Docking

Molecular simulations of the interaction between the NLS-NUPR1L and Impα3 were performed using AutoDock Vina 1.1.2 [36], on the basis of a protocol already used to screen the binding of the NLS and other fragments of the parent protein NUPR1 [12,60]. The structure of ΔImpα3 (without the IBB) was modelled starting from entry 5X8N of the Protein Data Bank (PDB), in which the intact monomeric protein is crystallized in complex with the NLS of the EBNA-LP protein of the Epstein–Barr virus [37].

The 24-residue (capped) sequence Ac-RTRREQALRTNWPAPGGHERKVAQ-NH_2 for the NLS-NUPR1L peptide used in our experiments possesses 99 rotatable dihedral angles; therefore, its conformational space is too large to be systematically explored. To overcame this difficulty, we employed nine 8-residue fragments (RTRREQAL, RREQALRT, GHERKVAQ) spanning the whole peptide sequence and differing by a shift of two consecutive amino acids. All the fragments were capped through acetylation (CH_3–CO–) and N-methyl amidation (–NH–CH_3) to mimic the missing regions of the peptide main chain, except the N-terminal end of the last fragment, in which standard amidation (–NH_2) was preserved. Docking simulations were carried out considering the whole protein surface (volume size 50 Å × 90 Å × 90 Å), and with very high exhaustiveness (up to 32 times larger than the default value) during the search [61].

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/21/19/7428/s1: There is one Table S1 containing the NMR assignment of the peptide, and four Figures in the Supplementary Material (Figures S1–S4).

Author Contributions: Conceptualization, J.L.N., B.R., A.V.-C., O.A. and J.L.I.; methodology, J.L.N., B.R., A.V.-C., O.A. and J.L.I.; investigation, J.L.N., B.R., A.V.-C. and A.J.-A.; data analysis, J.L.N., B.R., A.V.-C. and A.J.-A.; writing—original draft preparation, J.L.N., B.R. and A.V.-C.; writing—review and editing, J.L.N., B.R., J.L.I., A.J.-A., O.A. and A.V.-C.; funding acquisition, J.L.N., A.V.-C., O.A. and J.L.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spanish Ministry of Economy and Competitiveness and European ERDF Funds (MCIU/AEI/FEDER, EU) [RTI2018-097991-B-I00 to JLN and BFU2016-78232-P to AVC]; La Ligue Contre le Cancer, INCa, Canceropole PACA and INSERM to JLI; Miguel Servet Program from Instituto de Salud Carlos III [CPII13/00017 to OA]; Fondo de Investigaciones Sanitarias from Instituto de Salud Carlos III, and European Union (ERDF/ESF, 'Investing in your future') [PI15/00663 and PI18/00349 to OA]; Diputación General de Aragón [Protein Targets and Bioactive Compounds Group E45_17R to AVC, and Digestive Pathology Group B25_17R to OA]; and Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd).

Acknowledgments: We thank J. K. Forwood (Charles Sturt University, Waga Waga, Australia) for the kind gift of ΔImpα3 vector. B.R. acknowledges the kind hospitality and use of computational resources in the European Magnetic Resonance Center (CERM), Sesto Fiorentino (Florence), Italy. We thank the two anonymous reviewers for helpful suggestions and discussions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

ARM	Armadillo
CD	Circular dichroism
DOSY	Diffusion ordered spectroscopy
DIPSI	Decoupling in the presence of scalar interactions
IBB	Importin β-binding domain
IDP	Intrinsically disordered protein
Impα3	Human importin α3 isoform (residues 1–521)
ΔImpα3	Truncated species of Impα3 (residues 64-521) depleted of the IBB
ITC	Isothermal titration calorimetry
NLS	Nuclear localization sequence
NLS-NUPR1L	Nuclear localization sequence of NUPR1L (residues 51–74)
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NPC	Nuclear pore complex
NUPR1	Nuclear protein 1
NUPR1L	The NUPR1-like paralogue
TOCSY	Total correlation spectroscopy
TPPI	Time-proportional-phase incrementation technique
UV	Ultraviolet

References

1. Mallo, G.V.; Fiedler, F.; Calvo, E.L.; Ortiz, E.M.; Vasseur, S.; Keim, V.; Morisset, J.; Iovanna, J.L. Cloning and expression of the rat p8 cDNA, a new gene activated in pancreas during the acute phase of pancreatitis, pancreatic development, and regeneration, and which promotes cellular growth. *J. Biol. Chem.* **1997**, *72*, 32360–32369. [CrossRef]
2. Chowdury, U.R.; Samant, R.S.; Fodstat, O.; Shevde, L.A. Emerging role of nuclear protein 1 (nupr1) in cancer biology. *Cancer Metastasis Rev.* **2009**, *28*, 225–232. [CrossRef] [PubMed]
3. Uversky, V.N. A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci.* **2013**, *22*, 693–724. [CrossRef] [PubMed]
4. Berlow, R.B.; Dyson, H.J.; Wright, P.E. Expanding the paradigm: Intrinsically disordered proteins and allosterism. *J. Mol. Biol.* **2018**, *430*, 2309–2320. [CrossRef] [PubMed]
5. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Mol. Cell Biol.* **2015**, *16*, 18–29. [CrossRef] [PubMed]

6. Cano, C.E.; Hamidi, T.; Sandi, M.J.; Iovanna, J.L. Nupr-1: The Swiss knife of cancer. *J. Cell Physiol.* **2011**, *226*, 1439–1443. [CrossRef]
7. Goruppi, S.; Iovanna, J.L. Stress-inducible protein p8 is involved in several physiological and pathological processes. *J. Biol. Chem.* **2010**, *285*, 1577–1581. [CrossRef]
8. Hamidi, T.; Algül, H.; Cano, C.E.; Sandi, M.J.; Molejon, M.I.; Riemann, M.; Calvo, E.L.; Lomberk, G.; Dagorn, J.C.; Weih, F.; et al. Nuclear protein 1 promotes pancreatic cancer development and protects cells from stress by inhibiting apoptosis. *J. Clin. Investig.* **2012**, *122*, 2092–2103. [CrossRef]
9. Malicet, C.; Giroux, V.; Vasseur, S.; Dagorn, J.C.; Neira, J.L.; Iovanna, J.L. Regulation of apoptosis by the p8/prothymosin alpha complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2671–2676. [CrossRef]
10. Encinar, J.A.; Mallo, G.V.; Mizyrycki, C.; Giono, L.; González-Ros, J.M.; Rico, M.; Cánepa, E.; Moreno, S.; Neira, J.L.; Iovanna, J.L. Human p8 is a HMG-I/Y-like protein with DNA binding activity enhanced by phosphorylation. *J. Biol. Chem.* **2001**, *276*, 2742–2751. [CrossRef]
11. Aguado-Llera, D.; Hamidi, T.; Doménech, R.; Pantoja-Uceda, D.; Gironella, M.; Santoro, J.; Velázquez-Campoy, A.; Neira, J.L.; Iovanna, J.L. Deciphering the binding between Nupr1 and MSL1 and their DNA-repairing activity. *PLoS ONE* **2013**, *8*, e78101. [CrossRef]
12. Santofimia-Castaño, P.; Rizzuti, B.; Pey, A.L.; Soubeyran, P.; Vidal, M.; Urrutia, R.; Iovanna, J.L.; Neira, J.L. Intrinsically disordered chromatin protein NUPR1 binds to the C-terminal region of Polycomb RING1B. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6332–6341.
13. López, M.B.; García, M.N.; Grasso, D.; Bintz, J.; Molejón, M.I.; Vélez, G.; Lomberk, G.; Neira, J.L.; Urrutia, R.; Iovanna, J.L. Functional characterization of NUPR1L, a novel p53-regulated isoform of the high-mobility group (HMG)-related protumoral protein NUPR1. *J. Cell. Physiol.* **2015**, *230*, 2936–2950. [CrossRef] [PubMed]
14. Neira, J.L.; López, M.B.; Sevilla, P.; Rizzuti, B.; Cámara-Artigas, A.; Vidal, M.; Iovanna, J.L. The chromatin nuclear protein NUPR1L is intrinsically disordered and binds to the same proteins as its paralogue. *Biochem. J.* **2018**, *475*, 2271–2291. [CrossRef]
15. Stewart, M. Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell. Biol.* **2007**, *8*, 195–208. [CrossRef] [PubMed]
16. Bednenko, J.; Cingolari, G.; Gerace, L. Nucleo-cytoplasmic transport navigating the channel. *Traffic* **2003**, *4*, 127–135. [CrossRef]
17. Cingolani, G.; Bednenko, J.; Gillespie, M.T.; Gerace, L. Molecular basis for the recognition of a non-classical nuclear localization signal by importin beta. *Mol. Cell* **2002**, *10*, 1345–1353. [CrossRef]
18. Goldfarb, D.S.; Corbett, A.H.; Mason, D.A.; Harreman, M.T.; Adam, S.A. Importin: A multipurpose nuclear-transport receptor. *Trends Cell Biol.* **2004**, *14*, 505–514. [CrossRef]
19. Pumroy, R.A.; Cingolani, G. Diversification of importin-alpha isoforms in cellular trafficking and disease states. *Biochem. J.* **2015**, *466*, 13–28. [CrossRef]
20. Miyamoto, Y.; Loveland, K.L.; Yoneda, Y. Nuclear importin α and its physiological importance. *Comm. Integ. Biol.* **2012**, *5*, 220–222. [CrossRef]
21. Marvaldi, L.; Panayotokis, N.; Alber, S.; Dgan, S.Y.; Okladnikov, N.; Koppel, I.; Di Pizio, A.; Song, D.-A.; Tzur, Y.; Terenzio, M.; et al. Importin 3 regulates chronic pain pathways in peripheral sensory neurons. *Science* **2020**, *369*, 842–846. [CrossRef] [PubMed]
22. Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin. *Nat. Struct. Biol.* **1999**, *6*, 388–397. [CrossRef] [PubMed]
23. Lan, W.; Santofimia-Castaño, P.; Swayden, M.; Xia, Y.; Zhou, Z.; Audebert, S.; Camoin, L.; Huang, C.; Peng, L.; Jiménez-Alesanco, A.; et al. ZZW-115-dependent inhibition of NUPR1 nuclear translocation sensitizes cancer cells to genotoxic agents. *JCI Insight* **2020**, 138117. [CrossRef] [PubMed]
24. Valacco, M.P.; Varone, C.; Malicet, C.; Cánepa, E.; Iovanna, J.L.; Moreno, S. Cell growth-dependent subcellular localization of p8. *J. Cell Biochem.* **2006**, *97*, 1066–1079. [CrossRef] [PubMed]
25. Díaz-García, C.; Hornos, F.; Giudici, A.M.; Cámara-Artigas, A.; Luque-Ortega, J.L.; Arbe, A.; Rizzuti, B.; Alfonso, C.; Forwood, J.K.; Iovanna, J.L.; et al. Human importin α3 and its N-terminal truncated form, without the importin-β-binding domain, are oligomeric species with a low conformational stability in solution. *BBA Gen. Subj.* **2020**, *1864*, 129609. [CrossRef] [PubMed]
26. Pace, C.N.; Scholtz, J.M. Measuring the Conformational Stability of A Protein. In *Protein Structure*, 2nd ed.; Creighton, T.E., Ed.; Oxford University Press: Oxford, UK, 1997; pp. 253–259.
27. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, NY, USA, 1986.

28. Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ (1-40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magn. Reson. Chem.* **2002**, *40*, S89–S97. [CrossRef]
29. Kjaergaard, M.; Brander, S.; Poulsen, F.M. Random coil chemical shifts for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* **2011**, *49*, 139–149. [CrossRef]
30. Kjaergaard, M.; Poulsen, F.M. Sequence correction of random coil chemical shifts: Correlation between neighbour correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **2011**, *50*, 157–165. [CrossRef]
31. Cavanagh, J.; Fairbrother, W.J.; Palmer, A.G.; Skelton, N.J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, NY, USA, 1996.
32. Neira, J.L.; Hornos, F.; Cozza, C.; Cámara-Artigas, A.; Abián, O.; Velázquez-Campoy, A. The histidine phosphocarrier protein, HPr, binds to the highly thermostable regulator of sigma D protein, Rsd, and its isolated helical fragments. *Arch. Biochem. Biophys.* **2018**, *639*, 26–37. [CrossRef]
33. Cremades, N.; Velázquez-Campoy, A.; Freire, E.; Sancho, J. The flavodoxin from *Helicobacter pylori*: Structural determinants of thermostability and FMN cofactor binding. *Biochemistry* **2008**, *47*, 627–639. [CrossRef]
34. Bollen, Y.J.; Westphal, A.H.; Lindhoud, S.; van Berkel, W.J.; van Mierlo, C.P. Distant residues mediate picomolar binding affinity of a protein cofactor. *Nat. Commun.* **2012**, *3*, 1010. [CrossRef] [PubMed]
35. Yadahalli, S.; Neira, J.L.; Johnson, C.M.; Tan, Y.S.; Rowling, P.J.E.; Chattopadhyay, A.; Verma, C.S.; Itzhaki, L.S. Kinetic and thermodynamic effects of phosphorylation on p53 binding to MDM2. *Sci. Rep.* **2019**, *24*, 693. [CrossRef] [PubMed]
36. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef] [PubMed]
37. Nakada, R.; Matsuura, Y. Crystal structure of importin-alpha bound to the nuclear localization signal of Epstein-Barr virus EBNA-LP protein. *Protein Sci.* **2017**, *26*, 1231–1235. [CrossRef]
38. Delano, W.L. Available online: http://www.pymol.org/ (accessed on 5 October 2002).
39. Junod, S.L.; Kelich, J.M.; Ma, J.; Yang, W. Nucleocytoplasmic transport of intrinsically disordered proteins studies by high-speed super-resolution microscopy. *Protein Sci.* **2020**, *29*, 1459–1472. [CrossRef]
40. Smith, K.M.; Tsimbalyuk, S.; Edwards, M.G.; Cross, E.M.; Batra, J.; Soares da Costa, T.P.; Aragao, D.; Basler, C.F.; Forwood, J.K. Structural basis for importin alpha 3 specificity of W proteins in Hendra and Nipah viruses. *Nat. Commun.* **2018**, *9*, 3703. [CrossRef]
41. Neira, J.L.; Rizzuti, B.; Jiménez-Alesanco, A.; Palomino-Schätzlein, M.; Abián, O.; Velázquez-Campoy, A.; Iovanna, J.L. A Phosphorylation-Induced Switch in the Nuclear Localization Sequence of the Intrinsically Disordered NUPR1 Hampers Binding to Importin. *Biomolecules* **2020**, *10*, 1313. [CrossRef]
42. Miyatake, H.; Sanjoh, A.; Unzai, S.; Mtsuda, G.; Tatsumi, Y.; Miyamoto, Y.; Dohmae, N.; Aida, Y. Crystal structure of human importin α1 (Rch1) revealing a potential autoinhibition mode involving homodimerization. *PLoS ONE* **2015**, *20*, e0115995. [CrossRef]
43. Gill, S.C.; von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319–326. [CrossRef]
44. Kosugi, S.; Hasebe, M.; Tomita, M.; Yanagawa, H. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10171–10176. [CrossRef]
45. Kosugi, S.; Hasebe, M.; Matsumura, N.; Takashima, H.; Miyamoto-Sato, E.; Tomita, M.; Yanagawa, H. Six classes of nuclear localization signals specific to different binding grooves of importin α. *J. Biol. Chem.* **2009**, *284*, 478–485. [CrossRef]
46. Neira, J.L.; Hornos, F.; Bacarizo, J.; Cámara-Artigas, A.; Gómez, J. The monomeric species of the regulatory domain of tyrosine hydroxylase has a low conformational stability. *Biochemistry* **2017**, *55*, 3418–3431. [CrossRef] [PubMed]
47. Birdsall, B.; King, R.W.; Wheeler, M.R.; Lewis, C.A., Jr.; Goode, S.; Dunlap, R.B.; Roberts, G.C. Correction for light absorption in fluorescence studies of protein-ligand interactions. *Anal. Biochem.* **1983**, *132*, 353–361. [CrossRef]
48. Beckett, D. Measurement and analysis of equilibrium binding titrations: A beginner's guide. *Methods Enzymol.* **2011**, *488*, 1–16. [PubMed]

49. Royer, C.A.; Scarlatta, S.F. Fluorescence approaches to quantifying biomolecular interactions. *Methods Enzymol.* **2008**, *450*, 79–106. [PubMed]
50. Benjwal, S.; Verma, S.; Röhm, K.H.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* **2006**, *15*, 635–639. [CrossRef]
51. Santofimia-Castaño, P.; Xia, Y.; Lan, W.; Zhou, Z.; Huang, C.; Peng, L.; Soubeyran, P.; Velázquez-Campoy, A.; Abian, O.; Rizzuti, B.; et al. Ligand-based design identifies a potent NUPR1 inhibitor exerting anticancer activity via necroptosis. *J. Clin. Investig.* **2019**, *129*, 2500–2513.
52. Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992**, *2*, 661–675. [CrossRef]
53. Wilkins, D.K.; Grimshaw, S.B.; Receveur, V.; Dobson, C.M.; Jones, J.A.; Smith, L.J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR technique. *Biochemistry* **1999**, *38*, 16424–16431. [CrossRef]
54. Marion, D.; Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **1983**, *11*, 967–975. [CrossRef]
55. Bax, A.; Davis, D.G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **1985**, *65*, 355–360. [CrossRef]
56. Kumar, A.; Ernst, R.R.; Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6. [CrossRef]
57. Cavanagh, J.; Rance, M. Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. *J. Magn. Reson.* **1992**, *96*, 660–678. [CrossRef]
58. Anglister, J.; Grzesiek, S.; Ren, H.; Klee, C.B.; Bax, A. Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. *J. Biomol. NMR* **1993**, *3*, 121–126. [CrossRef]
59. Sklenar, V.; Bax, A. Spin echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* **1987**, *74*, 469–479. [CrossRef]
60. Neira, J.L.; Correa, J.; Rizzuti, B.; Santofimia-Castaño, P.; Abian, O.; Velázquez-Campoy, A.; Fernández-Megía, E.; Iovanna, J.L. Dendrimers as competitors of protein-protein interactions of the intrinsically disordered nuclear chromatin protein NUPR1. *Biomacromolecules* **2019**, *20*, 2567–2576. [CrossRef]
61. Grande, F.; Rizzuti, B.; Occhiuzzi, M.A.; Ioele, G.; Casacchia, T.; Gelmini, F.; Guzzi, R.; Garofalo, A.; Statti, G. Identification by molecular docking of homoisoflavones from *Leopoldia comosa* as ligands of estrogen receptors. *Molecules* **2018**, *23*, 894. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Annex II: Participation in other projects

1. Peralta-Moreno MN, Anton-Muñoz V, Ortega-Alarcon D, **Jimenez-Alesanco A**, Vega S, Abian O, et al. Autochthonous Peruvian Natural Plants as Potential SARS-CoV-2 Mpro Main Protease Inhibitors. Pharmaceuticals. 2023;16(4):1-18.

Investigation related to SARS-CoV-2 M^{pro} expression and purification, SARS-CoV-2 M^{pro} proteolytic activity and SARS-CoV-2 M^{pro} inhibition assays.

2. Bastos M, Abian O, Johnson CM, Ferreira-da-silva F, Vega S, **Jimenez-Alesanco A**, et al. Isothermal titration calorimetry. Nat Rev methods Prim. 2023;3(17).

Reproducibility and data deposition.

3. Rubio-Martinez J, **Jimenez-Alesanco A**, Ceballos-Laita L, Ortega-Alarcon D, Vega S, Calvo C, et al. Discovery of Diverse Natural Products as Inhibitors of SARS-CoV-2 MproProtease through Virtual Screening. J Chem Inf Model. 2021;61(12):6094-106.

Investigation related to SARS-CoV-2 M^{pro} expression and purification, SARS-CoV-2 M^{pro} proteolytic activity and SARS-CoV-2 M^{pro} inhibition assays.

4. Sancineto L, Ostacolo C, Ortega-Alarcon D, **Jimenez-Alesanco A**, Ceballos-Laita L, Vega S, et al. L-arginine improves solubility and anti SARS-CoV-2 Mpro activity of rutin but not the antiviral activity in cells. Molecules. 2021;26(19):1-14.

Methodology, formal analysis and investigation related to SARS-CoV-2 M^{pro} expression and purification, SARS-CoV-2 M^{pro} proteolytic activity and SARS-CoV-2 M^{pro} inhibition assays.

5. Rizzuti B, Ceballos-Laita L, Ortega-Alarcon D, **Jimenez-Alesanco A**, Vega S, Grande F, et al. Sub-micromolar inhibition of SARS-CoV-2 3CLpro by natural compounds. Pharmaceuticals. 2021;14(9):1-10.

Validation, formal analysis, investigation, visualization, writing-review and editing related to SARS-CoV-2 3CLpro expression and purification, SARS-CoV-2 3CLpro catalytic activity, SARS-CoV-2 3CLpro inhibition, circular dichroism and emission fluorescence, and PAGE native electrophoresis assays.

6. Rizzuti B, Grande F, Conforti F, **Jimenez-Alesanco A**, Ceballos-Laita L, Ortega-Alarcon D, et al. Rutin is a low micromolar inhibitor of SARS-CoV-2 main protease 3CLpro: Implications for drug design of quercetin analogs. Biomedicines. 2021;9(4):1-20.

Validation, formal analysis, investigation, visualization, writing-review and editing related to SARS-CoV-2 3CLpro expression and purification, circular dichroism and fluorescence spectroscopy, SARS-CoV-2 3CLpro catalytic activity, SARS-CoV-2 3CLpro inhibition and isothermal titration calorimetry assays.

7. Mangiavacchi F, Botwina P, Menichetti E, Bagnoli L, Rosati O, Marini F, et al. Seleno-functionalization of quercetin improves the non-covalent inhibition of Mpro and its antiviral activity in cells against SARS-CoV-2. Int J Mol Sci. 2021;22(13).

Methodology, formal analysis and investigation related to SARS-CoV-2 M^{pro} expression and purification, SARS-CoV-2 M^{pro} proteolytic activity and SARS-CoV-2 M^{pro} inhibition assasy.

8. Neira JL, **Jimenez-Alesanco A**, Rizzuti B, Velazquez-Campoy A. The nuclear localization sequence of the epigenetic factor RYBP binds to human importin α3. Biochim Biophys Acta - Proteins Proteomics. 2021;1869(8).

Investigation related to isothermal titration calorimetry assays.

9. Losada-Garcia N, **Jimenez-Alesanco A**, Velazquez-Campoy A, Abian O, Palomo JM. Enzyme/Nanocopper Hybrid Nanozymes: Modulating Enzyme-like Activity by the Protein Structure for Biosensing and Tumor Catalytic Therapy. ACS Appl Mater Interfaces. 2021;13(4):5111-24.

Investigation related to circular dichroism, fluorescence spectroscopy and cell viability assays.

10. Bensabeh N, **Jimenez-Alesanco A**, Liblikas I, Ronda JC, Cádiz V, Galia M, et al. Biosourced all-acrylic ABA block copolymers with lactic acid-based soft phase. Molecules. 2020;25(23).

Methodology, investigation, writing-review and editing related to cell viability assays.

11. Moreno A, **Jimenez-Alesanco A**, Ronda JC, Cadiz V, Galia M, Percec V, et al. Dual Biochemically Breakable Drug Carriers from Programmed Telechelic Homopolymers. Biomacromolecules. 2020;21(10):4313-25.

Methodology, investigation, formal analysis, writing, writing-review and editing related to cell viability and cell uptake studies.

12. Abian O, Ortega-Alarcon D*, **Jimenez-Alesanco A***, Ceballos-Laita L*, Vega S, Reyburn HT, et al. Structural stability of SARS-CoV-2 3CLpro and identification of

quercetin as an inhibitor by experimental screening. Vol. 164, International Journal of Biological Macromolecules. 2020. p. 1693-703. (* Equal contribution).

Validation, formal analysis, investigation, visualization, writing-review and editing related to SARS-CoV-2 3CLpro expression and purification, circular dichroism and fluorescence spectroscopy, dynamic light scattering, size-exclusion chromatography, SARS-CoV-2 3CLpro proteolytic activity, activity-based screening, SARS-CoV-2 3CLpro inhibition, thermal shift assay and isothermal titration calorimetry assays.

In addition, studies related to the inhibition of SARS-CoV-2 3CLpro resulted in the following patent:

Title: Coating agent based on a biohybrid of copper nanoparticles and its use as a biocidal agent. Authors: Palomo-Carmona J.M., Abian, O., Velazquez, A., Ortega-Alarcon, D., **Jimenez-Alesanco, A.**, Ceballos Laita, L. Registration number: ES1641.1627. Application number: P202031282. Date of submission: December 22, 2020. Presentation site: OEPM Madrid. Entity: CSIC.

Annex III: Supplementary tables

Table S1. Approaches to ETBF infection diagnosis ([a] Parental HT-29 cells sometimes substituted; [b] Usually lamb or rabbit; [c] RITARD: reversible ileal-tie adult rabbit diarrhea) (42,55,56).

Diagnostic method	Advantages	Limitations
Stool culture from rectal swabs or bulk stool collection	*B. fragilis* isolation for direct confirmation of: - Genes by PCR/RT-PCR - BFT secretion by HT-29/C1 assay*	It can be laborious and costly, and it may cause delays in the diagnosis because it depends on expertise in anaerobic microbiology.
Isolation from blood or tissue for PCR	It is a rapid testing for the presence or absence of *B. fragilis.*	This technique needs broth amplification or tissue homogenization, which involve several steps. It can be expensive and labor-intensive, requiring anaerobic microbiology expertise and specialized equipment.
PCR from stool	It provides quicker results compared to stool culture for *B. fragilis.*	It requires the extraction of fecal DNA, which needs more steps. The sensitivity could be potentially limited by fecal inhibitors of PCR. However, using overnight enrichment cultures, the sensitivity can be enhanced.
HT-29/C1 cell assay [a]	The detection of the BFT biological activity can be obtained directly in stool or in culture supernatants of *B. fragilis* isolates. This assay shows excellent correlation with lamb intestinal loop assay. This assay detects, with 100% of specificity, as low as 0.5 pM BFT comparing to lamb ileal loop assay. This assay is an i*n vitro* method.	It can be expensive and labor-intensive, requiring anaerobic microbiology expertise and specialized equipment. The interpretation can be subjective, because the diagnosis is by observing microscopically the cellular morphological changes caused by BFT (Figure 12, main text). This could be more objective using BFT-neutralizing antibodies.
Enzyme-linked immunosorbent assay for fecal BFT	It is potentially a rapid diagnostic approach.	There is only limited data supporting the detection of BFT in stool.
Immunomagnetic separation polymerase chain reaction (IMS-PCR)	It is potentially a time-saving approach. Both *B. fragilis* isolation and gene detection by PCR can be performed in parallel on the same sample.	It requires noncommercial reagents, which could introduce variable performance to date.
Intestinal loop assays [b]	This approach detects the secretion stimulated by ETBF or BFT (accumulation of fluid in the lamb ileal loop).	It can be expensive and labor-intensive, and this approach requires a prior *B. fragilis* isolation (*in vivo* method). It requires the injection of isolated strains into sutured intestinal loops to measure subsequent fluid accumulation as a result of toxin response.
RITARD model [c]	This technique detects ETBF disease.	Requires a prior *B. fragilis* isolation. It is an expensive and labor-intensive method because it requires the injection of isolated strains into sutured intestinal loops to measure subsequent fluid accumulation as a result of toxin response (*in vivo* method).

Table S2. The amino acid sequences of BFT-1, BFT-2 and BFT-3 isoforms.

Isoform	Sequence
BFT-1	>tr|Q9S5W0|Q9S5W0_BACFG Fragilysin OS=Bacteroides fragilis OX=817 GN=bft-1 PE=4 SV=1 MKNVKLLLMLGTAALLAACSNEADSLTTSIDAPVTASIDLQSVSYTDLATQLNDVSDFGK MIILKDNGFNRQVHVSMDKRTKIQLDNENVRLFNGRDKDSTSFILGDEFAVLRFYRNGES ISYIAYKEAQMMNEIAEFYAAPFKKTRAINEKEAFECIYDSRTRSAGKDIVSVKINIDKA KKILNLPECDYINDYIKTPQVPHGITESQTRAVPSEPKTVYVICLRENGSTIYPNEVSAQ MQDAANSVYAVHGLKRYVNFHFVLYTTEYSCPSGDAKEGLEGFTASLKSNPKAEGYDDQI YFLIRWGTWDNKILGMSWFNSYNVNTASDFEASGMSTTQLMYPGVMAHELGHILGAEHTD NSKDLMYATFTGYLSHLSEKNMDIIAKNLGWEAADGD
BFT-2	>tr|O05091|O05091_BACFG Metalloprotease OS=Bacteroides fragilis OX=817 GN=bft-2 PE=4 SV=1 MKNVKLLLMLGTAALLAACSNEADSLTTSIDTPVTASIDLQSVSYTDLATQLNDVSDFGK MIILKDNGFNRQVHVSMDKRTKIQLDNENVRLFNGRDKDSTSFILGDEFAVLRFYRNGES ISYIAYKEAQMMNEIAEFYAAPFKKTRAINEKEAFECIYDSRTRSAGKDLVSVKINIDKA KKILNLPECDYINDYIKTPQVPHGITESQTRAVPSEPKTVYVICLRESGSTVYPNEVSAQ MQDAANSVYAVHGLKRFVNLHFVLYTTEYSCPSGNADEGLDGFTASLKANPKAEGYDDQI YFLIRWGTWDNNILGISWLDSYNVNTASDFKASGMSTTQLMYPGVMAHELGHILGARHAD DPKDLMYSKYTGYLFHLSEENMYRIAKNLGWEIADGD
BFT-3	>tr|O86049|O86049_BACFG BFT-3 OS=Bacteroides fragilis OX=817 GN=bft-3 PE=1 SV=2 MKNVKLLLMLGTAALLAACSNEADSLTTSIDAPVTASIDLQSVSYTDLATQLNDVSDFGK MIILKDNGFNRQVHVSMDKRTKIQLDNENVRLFNGRDKDSTNFILGDEFAVLRFYRNGES ISYIAYKEAQMMNEIAEFYAAPFKKTRAINEKEAFECIYDSRTRSAGKYPVSVKINVDKA KKILNLPECDYINDYIKTPQVPHGITESQTRAVPSEPKTVYVICLRENGSTVYPNEVSAQ MQDAANSVYAVHGLKRYVNLHFVLYTTEYACPSGNADEGLDGFTASLKANPKAEGYDDQI YFLIRWGTWDNNILGISWLNSYNVNTASDFKASGMSTTQLMYPGVMAHELGHILGANHAD DPKDLMYSKYTGYLFHLSEKNMDIIAKNLGWEIADGD

Table S3. Examples of several studies using *G. mellonella* in the microbiology field.

Bacterial species	Type of study	Reference
Acinetobacter baumannii	Comparison of virulence of different strains. Study of mutant strains for important virulence factors. Discovery of new hypervirulent strains. Host-pathogen interactions. Effects of antibacterial compounds.	(401–404)
Francisella tularensis	Effects of antibacterial compounds.	(405)
Pseudomonas aeruginosa	Identification of virulence factors. Discovery of novel components involved in host innate immune responses. Comparison of virulence of different strains. Screening for virulence of new strains.	(406–409)
Yersinia pseudotuberculosis	Study of mutant strains for important virulence factors.	(410)
Staphylococcus aureus	Effects of antibacterial compounds. Combination of antimicrobial therapies to treat infections. Control of infections with bacteriophage therapy.	(411–413)
Streptococcus pyogenes	Molecular basis of host-pathogen interactions.	(414)
Streptococcus mutans	Comparison of virulence of different strains. Study of mutant strains for important virulence factors.	(415)
Enterococcus faecalis	Study of mutant strains for important virulence factors. Identification of virulence factors.	(416,417)
Candida albicans	Identification of virulence factors. Comparison of virulence of different strains.	(418)
Cryptococcus neoformans	Identification of virulence factors. Host immune responses to infection. Effects of antifungal compounds.	(419)
Escherichia coli	Combination of antimicrobial therapies. Control of infections with bacteriophage therapy. Study of mutant strains for important virulence factors.	(420–422)
Klebsiella pneumoniae	Study of mutant strains for important virulence factors. Comparison of virulence of different strains.	(423,424)
Actinobacillus pleuropneumoniae	Discriminate the virulence of different isolates. Study of mutant strains.	(168,425)

Table S4. Crystallographic data.

Dataset	**proBFT-3** ***Ligand unbound***	**proBFT-3** ***MOA4 complex (TETR)***	**proBFT-3** ***MOA9 complex (ORTH)***	**proBFT-3** ***MOA9 complex (TETR)***	**proBFT-3** ***MOA10 complex (TETR)***
Beam line (synchrotron)	I04-1 (DIAMOND)	XALOC (ALBA)	XALOC (ALBA)	I04-1 (DIAMOND)	I04-1 (DIAMOND)
Space group / protomers per a.u. [a]	$P2_12_12_1$ / 2	$P4_12_12$ / 2	$P2_12_12_1$ / 2	$P4_12_12$ / 2	$P4_12_12$ / 2
Cell constants (a, b, c, in Å)	68.72, 82.74, 158.90	84.02, 84.02, 267.27	69.55, 83.06, 157.48	83.83, 83.83 266.32	83.95, 83.95, 266.28
Wavelength (Å)	0.91190	0.97926	0.97926	0. 91190	0.91190
Measurements / unique reflections	1,016,956 / 78,068	722,555 / 70.841	145,122 / 25,344	2.147.093 / 82,100	1,654,667 / 62,673
Resolution range (Å) (outermost shell) [b]	79.5 – 1.85 (1.96 – 1.85)	80.2 – 1.95 (2.07 – 1.95)	157.5 – 2.70 (2.86 – 2.70)	80.0 – 1.85 (1.96 – 1.85)	80.0 – 2.03 (2.15 – 2.03)
Completeness (%) / R_{merge} [c]	100 (99.9) / 0.129 (1.602)	100 (99.9) / 0.143 (1.345)	98.1 (98.3) / 0.184 (0.962)	100 (99.9) / 0.129 (1.737)	100 (99.9) / 0.115 (2.073)
R_{meas} [d] / CC($^1/_2$) [d]	0.134 (1.678) / 0.999 (0.706)	0.151 (1.416) / 0.999 (0.749)	0.202 (1.100) / 0.990 (0.730)	0.132 (1.774) / 1.0 (0.787)	0.118 (2.113) / 1.0 (0.830)
Average intensity [e]	13.4 (1.5)	12.2 (1.6)	9.8 (1.7)	18.9 (1.8)	23.7 (2.0)
B-Factor (Wilson) (Å^2) / Aver. multiplicity	35.4 / 13.0 (11.3)	38.2 / 10.2 (10.0)	56.3 / 5.7 (4.1)	36.6 / 26.2 (24.3)	44.2 / 26.4 (27.3)
Resolution range used for refinement (Å)	63.1 – 1.85	80.2 – 1.95	73.5 – 2.70	80.0 – 1.85	80.1 – 2.03
Reflections used (test set)	78.068 (748)	70,841 (716)	25.317 (706)	82,100 (743)	62,673 (747)
Crystallographic R_{factor} (free R_{factor}) [c]	0.189 (0.211)	0.201 (0.215)	0.216 (0.243)	0.191 (0.204)	0.211 (0.223)
Non-H protein atoms / ligands / waters per a.u.	5632 / 2 Zn^{2+}, 1 Mg^{2+}, 6 FOR, 7 PRO, 7 DMSO / 727	5393 / 1 Zn^{2+}, 2 Cl$^-$, 4 I04, 3 PRO, 6 DMSO / 591	5515 / 3 Zn^{2+}, 2 I04, 1 PGE 2 PRO, 1 ACT / 95	5525 / 2 Zn^{2+}, 3 Cl$^-$, 2 I09, 1 PGE, 1PEG, 4 DMSO, 5 EDO, 2 PRO / 777	5564 / 2 Zn^{2+}, 2 I10, 4 DMSO, 7 PRO / 665
Rmsd from target values bonds (Å) / angles (°)	0.008 / 0.91	0.008 / 0.89	0.009 / 0.92	0.008 / 0.89	0.008 / 0.90
Average B-factor (Å^2)	37.1	41.1	68.0	39.1	45.3
Protein contacts and geometry analysis [f]					
Ramachandran favoured/outliers/all	689 (98.4%) / 2 (0.3%) / 700	661 (98.9%) / 0 (0%) / 668	658 (96.8%) / 2 (0.3%) / 680	682 (97.7%) / 2 (0.3%) / 698	683 (98.4%) / 0 (0%) / 694
Bond-length/bond-angle/chiral./plan. outliers	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0	0 / 0 / 0 / 0
Side-chain outliers	9 (1.5%)	8 (1.4%)	27 (4.5%)	13 (2.1%)	6 (1.0%)
All-atom clashes / clashscore [f]	30 / 2.5	21 / 1.8	41 / 3.7	23 / 1.9	27 / 2.3
RSRZ outliers [f] / F_o:F_c correlation work (free)	42 (6.0%) / 0.956 (0.947)	35 (5.2%) / 0.945 (0.944)	28 (4.1%) / 0.911 (0.895)	43 (6.2%) / 0.950 (0.957)	34 (4.9%) / 0.932 (0.937)
PDB access code	7PND	7POL	7POO	7POQ	7POU

[a] Abbreviations: ACT, acetate; DMSO, dimethyl sulfoxide; EDO, ethylene glycol; FOR, formate anion; I04, compound MOA4; I09, compound MOA9; I10, compound MOA10; PEG, diethylene glycol; PGE, triethylene glycol; PRO, L-proline; RSRZ, real-space R-value Z-score.

[b] Values in parenthesis refer to the outermost resolution shell.

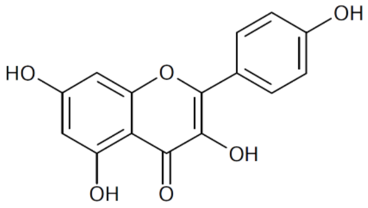
[c] For definitions, see Table 1 in (426).

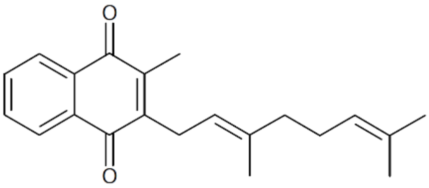
[d] For definitions, see (427,428).

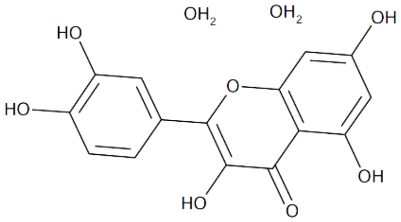
[e] Average intensity is <I/σ(I)> of unique reflections after merging according to *Xscale* (194).

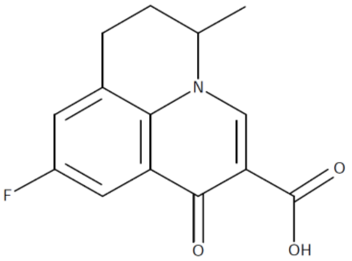
[f] According to the wwPDB Validation Service (https://wwpdb-validation.wwpdb.org/validservice).

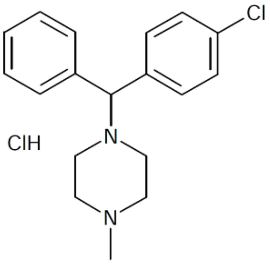
Table S5. MOA compound information.

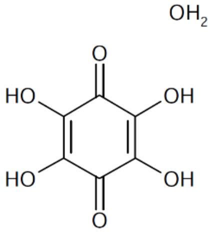
Compound		Chemical structure	MW (g/mol)
MOA1	**Kaempferol**		286.24
		Therapeutic information	
		Kaempferol is a natural plant product, common in vegetables, fruits, plants and herbal medicines. it reduces cancer, arteriosclerosis, cardiovascular disorders, and serves as an antioxidant and anti-inflammatory. Kaempferol modulates a number of key elements in the cellular signal transduction pathway linked to apoptosis, angiogenesis, inflammation, and metastasis.	

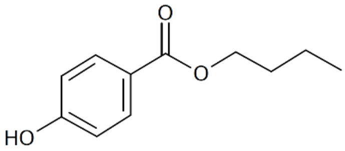
Compound		Chemical structure	MW (g/mol)
MOA2	**Vitamin K2**		308.42
		Therapeutic information	
		Vitamin K2 shows promise as a supplement in the treatment of osteoporosis (used in combination with vitamin D and calcium). It has demonstrated the ability to reduce vascular calcification, and holds potential in the fields of diabetes, cancer, and osteoarthritis.	

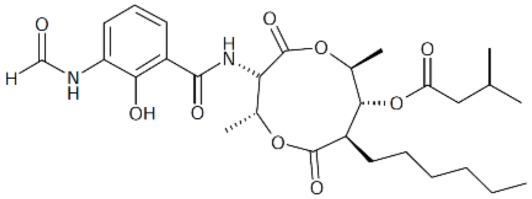
Compound		Chemical structure	MW (g/mol)
MOA3	**Quercetine dihydrate**		338.26
		Therapeutic information	
		Quercetin is a flavonoid found in many plants and foods, such as red wine, onions, green tea, apples, and berries. Quercetin has antioxidant and anti-inflammatory effects that might help reduce swelling, kill cancer cells, control blood sugar, and help prevent heart disease.	

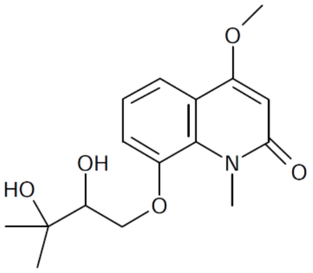
Compound		Chemical structure	MW (g/mol)
MOA4	**Flumequine**		261.25
		Therapeutic information	
		A synthetic chemotherapeutic antibiotic of the fluoroquinolone drug class used to treat bacterial infections. It kills bacteria by interfering with the enzymes that cause DNA to unwind and duplicate (DNA gyrase, topoisomerase IV). Flumequine was used, only in a limited number of countries, in veterinarian medicine for the treatment of enteric infections, as well as to treat cattle, swine, chickens, and fish.	

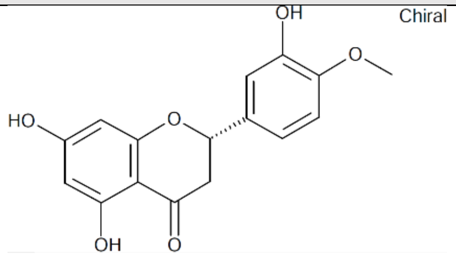
Compound		**Chemical structure**	**MW (g/mol)**
MOA5	**Chlorcyclizine hydrochloride**		337.29
		Therapeutic information	
		An antiemetic, antihistaminic and sedative compound, which acts targeting the histaminergic H1 receptor.	

Compound		**Chemical structure**	**MW (g/mol)**
MOA6	**Tetrahydroxy-1,4-quinone monohydrate**		190.11
		Therapeutic information	
		An anticataract agent, which is a redox active benzoquinone and it can take part in a redox cycle with semiquinone radicals, leading to the formation of ROS. So, it is an efficient inducer of ROS production. Tetrahydroxyquinone efficiently activates caspase-3, stimulates DNA fragmentation and provoke phosphatidylserine exposure. It also induces the release of cytochrome C (Cyt*c*) from the mitochondria.	

Compound		Chemical structure	MW (g/mol)
MOA7	**Butylparaben**		194.23
		Therapeutic information	
		Butylparaben is a chemical compound commonly used as an antifungal preservative in cosmetic products. It promotes apoptosis in human trophoblast cells through increased oxidative stress-induced endoplasmic reticulum stress (ERS).	

Compound		Chemical structure	MW (g/mol)
MOA8	**Antimycin A**	 O H N H OH O H N O O O O O O	548.64
		Therapeutic information	
		Antimycin A is an inhibitor of electron transport from cytochrome b to cytochrome complex III (Cyt*c* reductase). Binding of Antimycin A to Cyt*c* reductase inhibits the oxidation of ubiquinol, disrupts the Q-cycle of enzyme turnover, and halts cellular respiration.	

Compound		Chemical structure	MW (g/mol)
MOA9	**Foliosidine**		307.35
		Therapeutic information	
		A quinoline alkaloid, from plants of the Rutaceae family. Anticonvulsant, causes hypothermia, prevents cardiac arrhythmia caused by electric stimulation of the atria or by aconitine.	

Compound		Chemical structure	MW (g/mol)
MOA10	**Hesperetin**		302.3
		Therapeutic information	
		Hesperetin is a cholesterol lowering flavanoid found in a number of citrus juices. It appears to reduce cholesteryl ester mass and inhibit apoB secretion by up to 80%. Hesperetin may have antioxidant, anti-inflammatory, anti-allergic, hypolipidemic, vasoprotective and anticarcinogenic actions.	

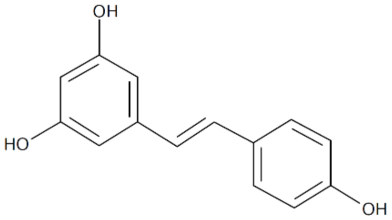
<table>
<tr><th colspan="2">Compound</th><th>Chemical structure</th><th>MW (g/mol)</th></tr>
<tr><td rowspan="3">MOA11</td><td rowspan="3">Resveratrol</td><td></td><td>228.24</td></tr>
<tr><th colspan="2">Therapeutic information</th></tr>
<tr><td colspan="2">Resveratrol is a chemical mostly found in red grapes. It has been shown to mimic effects of caloric restriction, exert anti-inflammatory and anti-oxidative effects, and affect the initiation and progression of many diseases through several mechanisms.</td></tr>
</table>

Table S6. The canonical sequence (isoform β) of NUPR1.

	Sequence
NUPR1	>sp|O60356|NUPR1_HUMAN Nuclear protein 1 OS=Homo sapiens OX=9606 GN=NUPR1 PE=1 SV=1 MATFPPATSAPQQPPGPEDEDSSLDESDLYSLAHSYLGGGGRKGRTKREAAANTNRPS PGGHERKLVTKLQNSERKKRGARR

Table S7. The sequence of NURP1L.

	Sequence
NUPR1L	>NP_001139184.1 nuclear protein 2 [Homo sapiens] MEAPAERALPRLQALARPPPPISYEEELYDCLDYYYLRDFPACGAGRSKGRTRREQAL RTNWPAPGGHERKVAQKLLNGQRKRRQRQLHPKMRTRLT

Table S8. The sequence of Impα3.

	Sequence
Impα3	>sp|O00629|IMA3_HUMAN Importin subunit alpha-3 OS=Homo sapiens OX=9606 GN=KPNA4 PE=1 SV=1 MADNEKLDNQRLKNFKNKGRDLETMRRQRNEVVVELRKNKRDEHLLKRRNVPHEDICE DSDIDGDYRVQNTSLEAIVQNASSDNQGIQLSAVQAARKLLSSDRNPPIDDLIKSGIL PILVHCLERDDNPSLQFEAAWALTNIASGTSEQTQAVVQSNAVPLFLRLLHSPHQNVC EQAVWALGNIIGDGPQCRDYVISLGVVKPLLSFISPSIPITFLRNVTWVMVNLCRHKD PPPPMETIQEILPALCVLIHHTDVNILVDTVWALSYLTDAGNEQIQMVIDSGIVPHLV PLLSHQEVKVQTAALRAVGNIVTGTDEQTQVVLNCDALSHFPALLTHPKEKINKEAVW FLSNITAGNQQQVQAVIDANLVPMIIHLLDKGDFGTQKEAAWAISNLTISGRKDQVAY LIQQNVIPPFCNLLTVKDAQVVQVVLDGLSNILKMAEDEAETIGNLIEECGGLEKIEQ LQNHENEDIYKLAYEIIDQFFSSDDIDEDPSLVPEAIQGGTFGFNSSANVPTEGFQF

Annex IV: Supplementary figures

Figure S1. International MIC distribution based on aggregated distributions of (A) MTZ and (B) FOX in *B. fragilis*. MIC distributions include collated data from multiple sources, geographical areas and time periods, and can never be used to infer rates of resistance (Figure from database: MIC EUCAST).

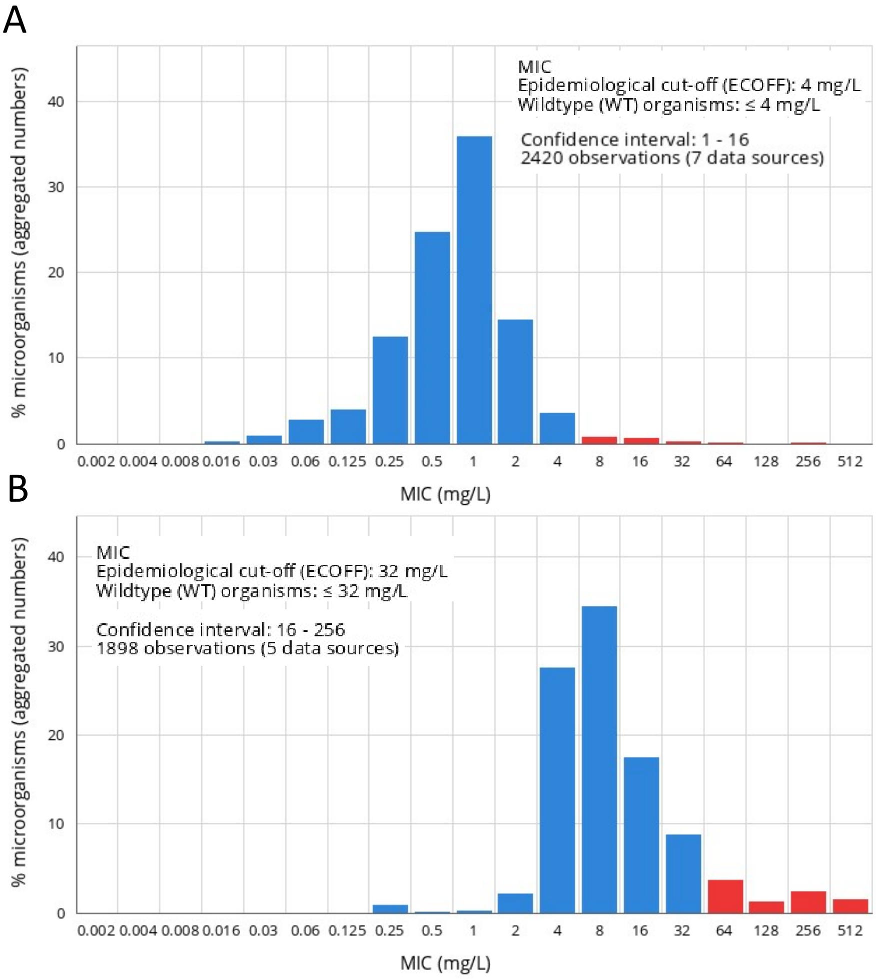


Figure S2. 24 h monitoring of the intrinsic cytotoxic effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae.

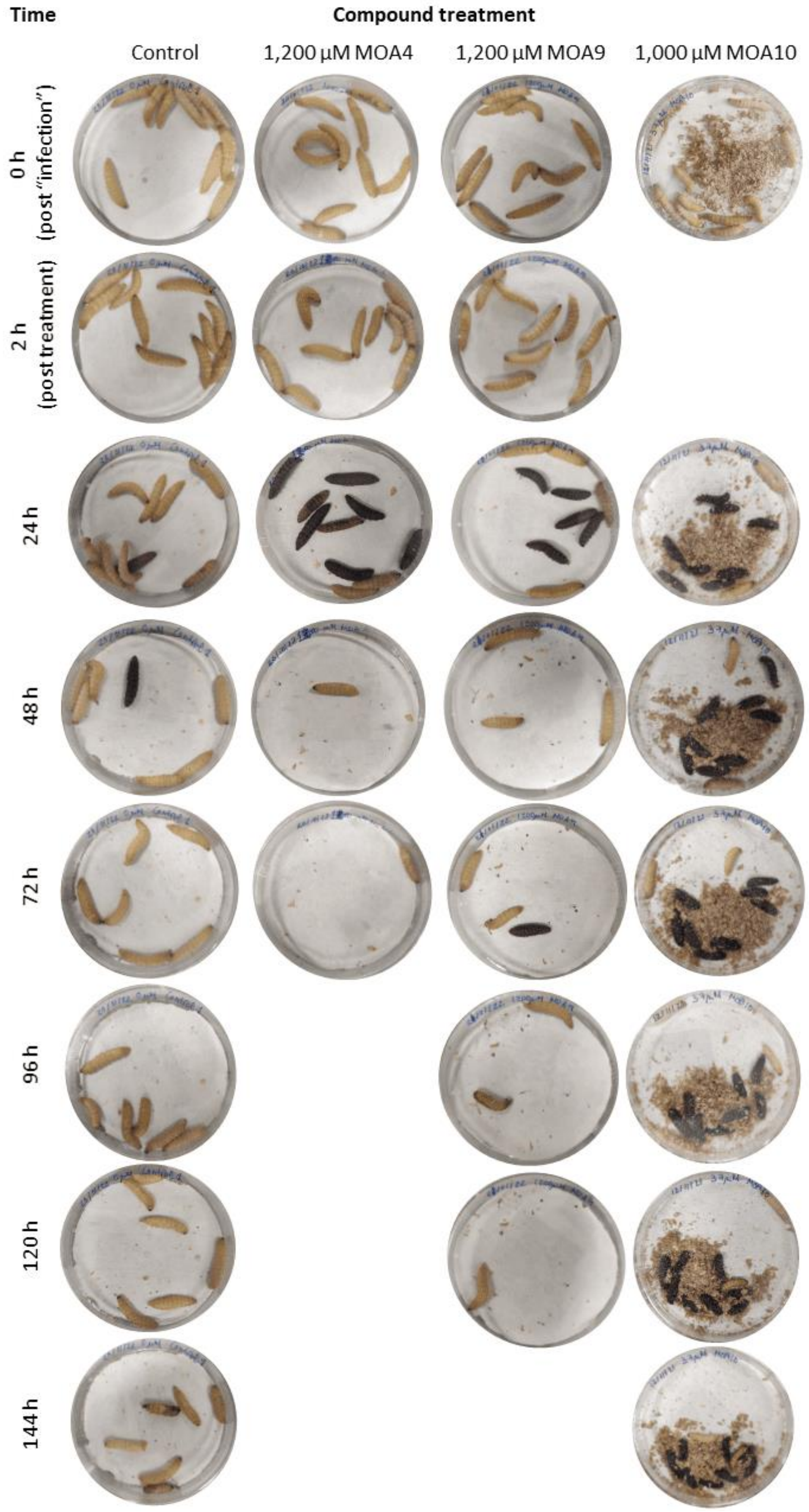

Time
Compound treatment
Control
1,200 µM MOA4
1,200 µM MOA9
1,000 µM MOA10
0 h (post "infection")
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S3. 24 h monitoring of the intrinsic cytotoxic effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae.

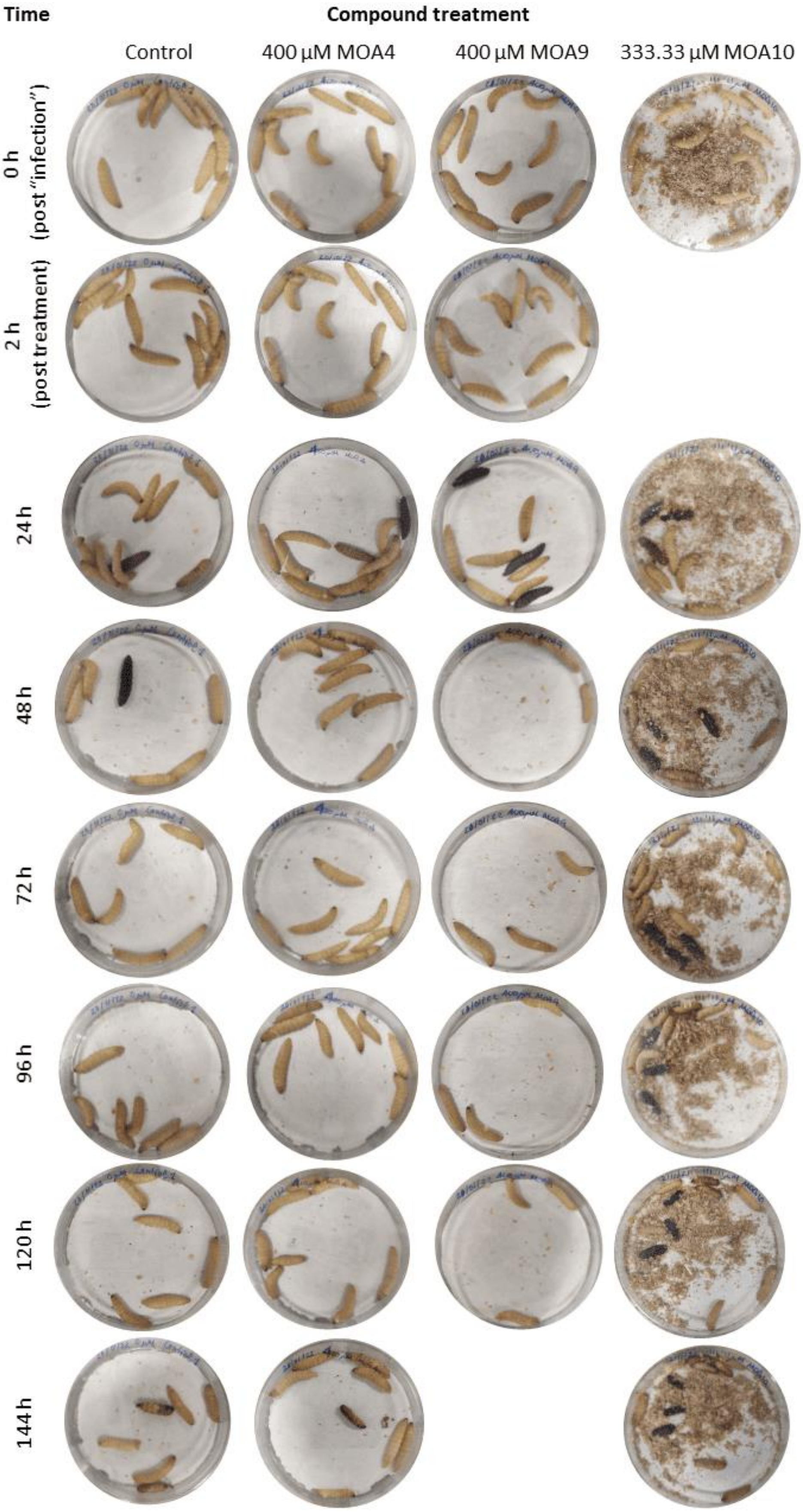

Time
Compound treatment
Control
400 µM MOA4
400 µM MOA9
333.33 µM MOA10
0 h (post "infection")
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S4. 24 h monitoring of the intrinsic cytotoxic effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae.

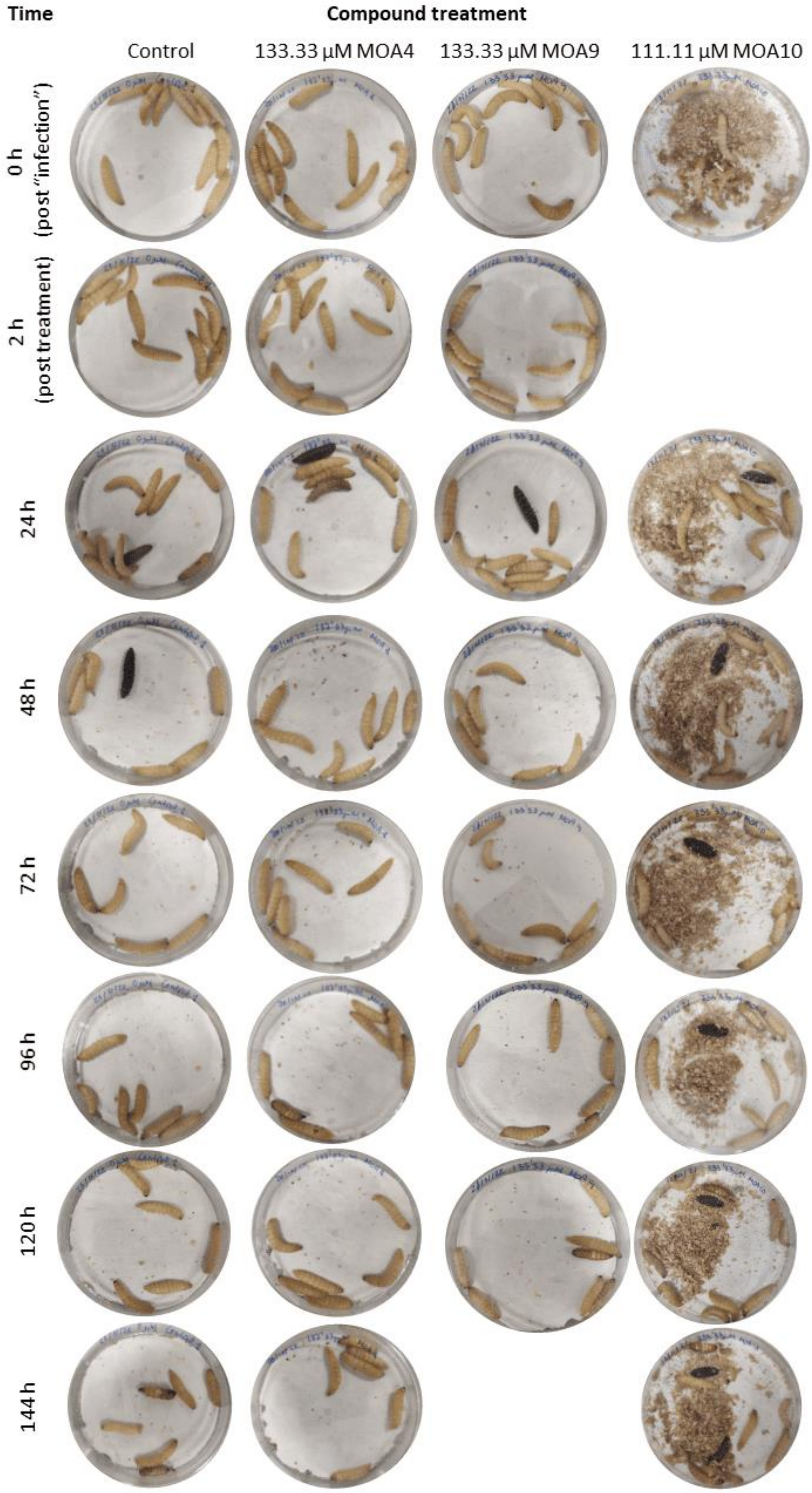

Time
Compound treatment
Control
133.33 μM MOA4
133.33 μM MOA9
111.11 μM MOA10
0 h (post "infection")
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S5. 24 h monitoring of the intrinsic cytotoxic effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae.

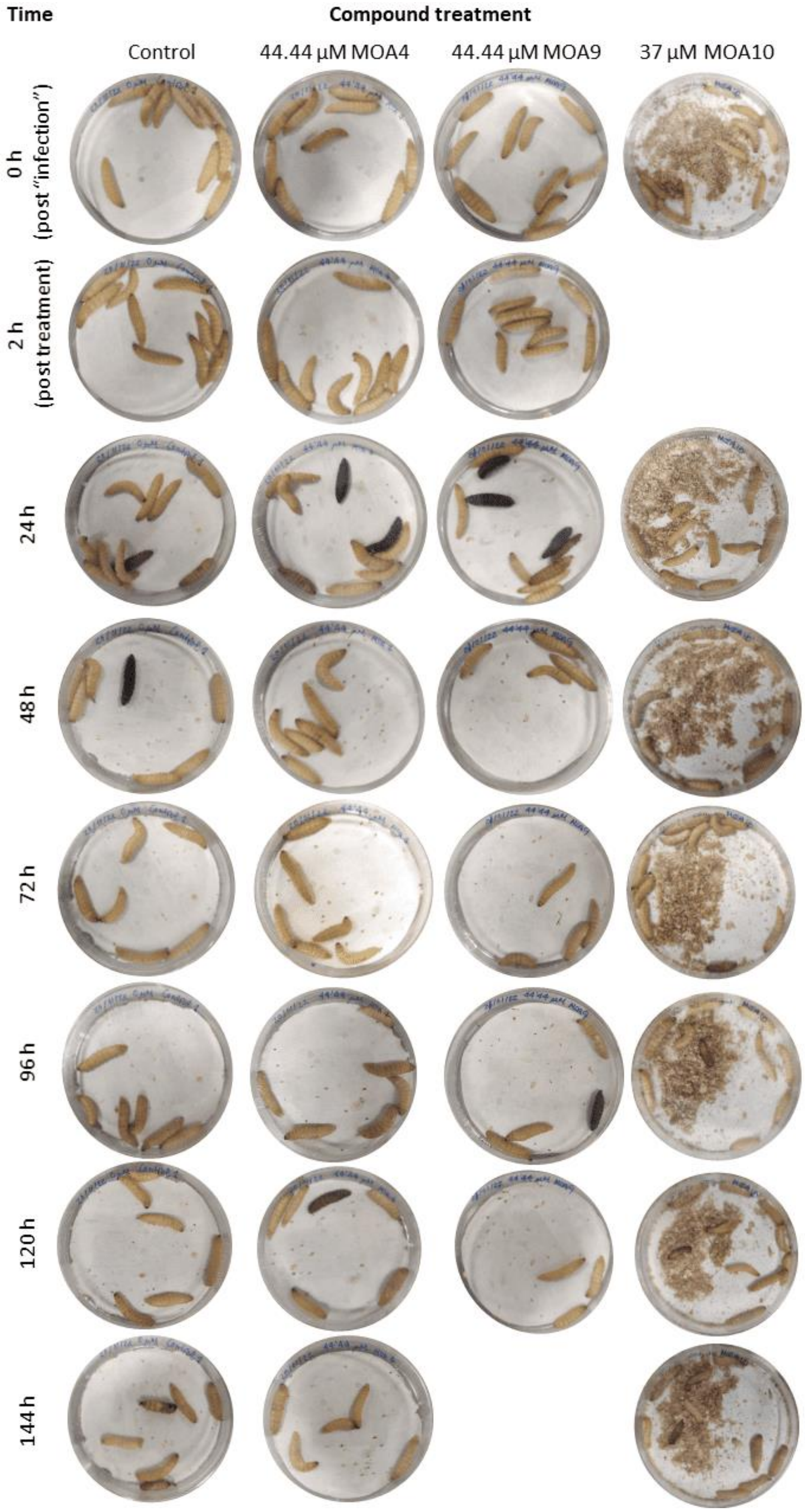

Time
Compound treatment
Control
44.44 µM MOA4
44.44 µM MOA9
37 µM MOA10
0 h (post "infection")
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S6. 24 h monitoring for determining the CFU/mL of *B. fragilis* necessary to produce a clear and rapid infection in *G. mellonella* larvae.

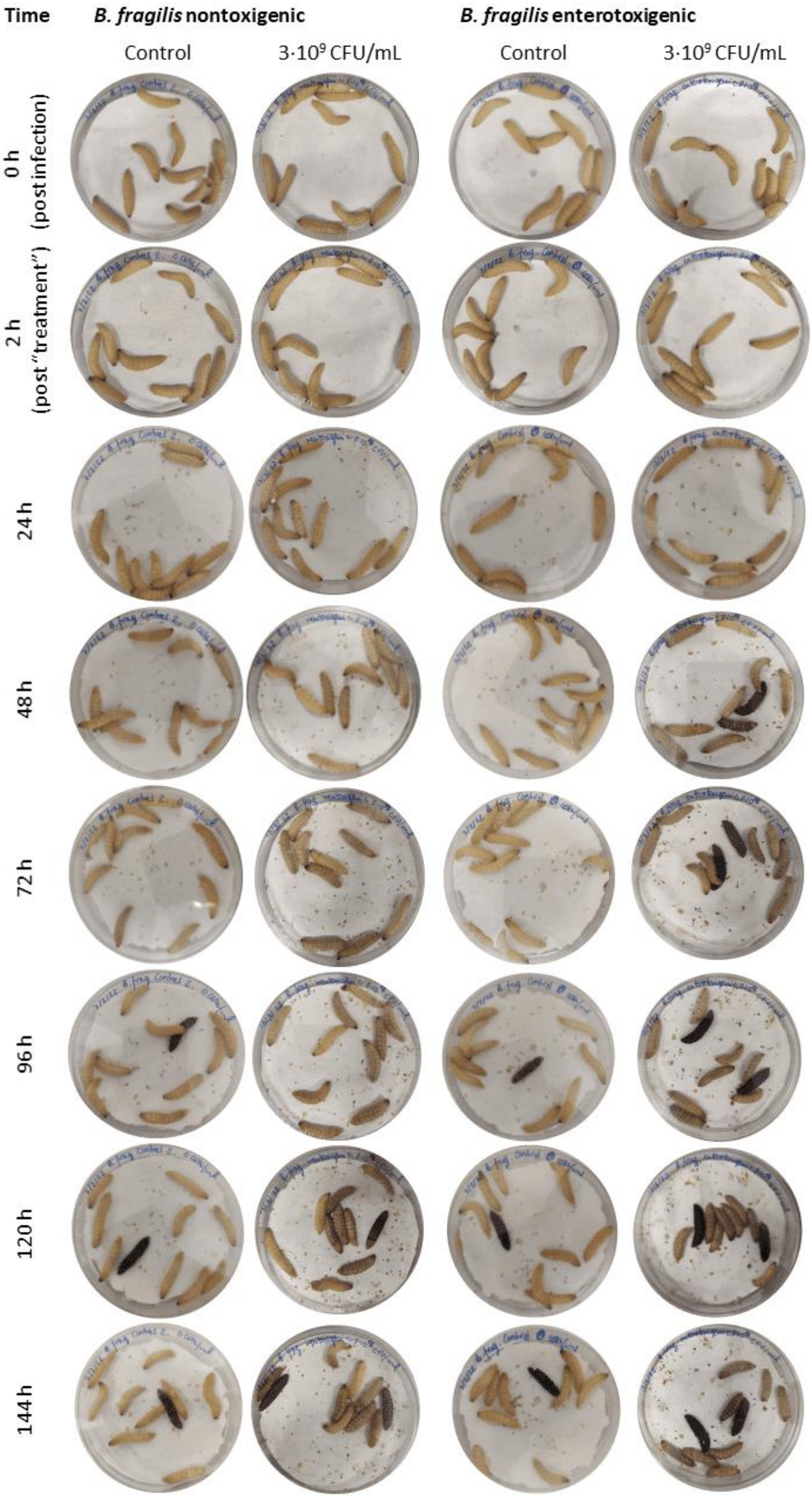

Time
B. fragilis nontoxigenic
B. fragilis enterotoxigenic
Control
$3\cdot10^9$ CFU/mL
Control
$3\cdot10^9$ CFU/mL
0 h (post infection)
2 h (post "treatment")
24 h
48 h
72 h
96 h
120 h
144 h

Figure S7. 24 h monitoring for determining the CFU/mL of *B. fragilis* necessary to produce a clear and rapid infection in *G. mellonella* larvae.

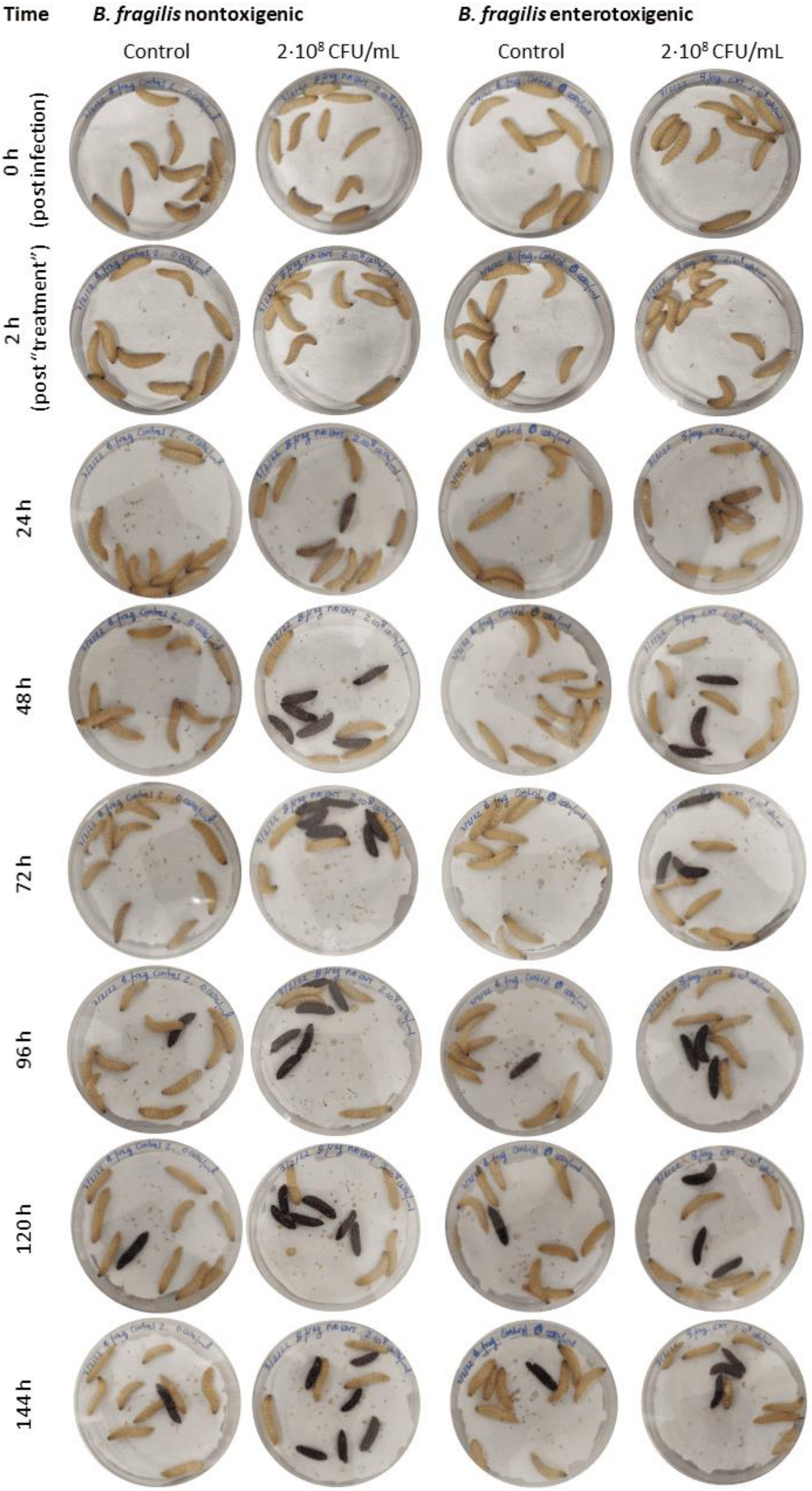

Time
B. fragilis nontoxigenic
B. fragilis enterotoxigenic
Control
2·10⁸ CFU/mL
Control
2·10⁸ CFU/mL
0 h (post infection)
2 h (post "treatment")
24 h
48 h
72 h
96 h
120 h
144 h

Figure S8. 24 h monitoring for determining the CFU/mL of *B. fragilis* necessary to produce a clear and rapid infection in *G. mellonella* larvae.

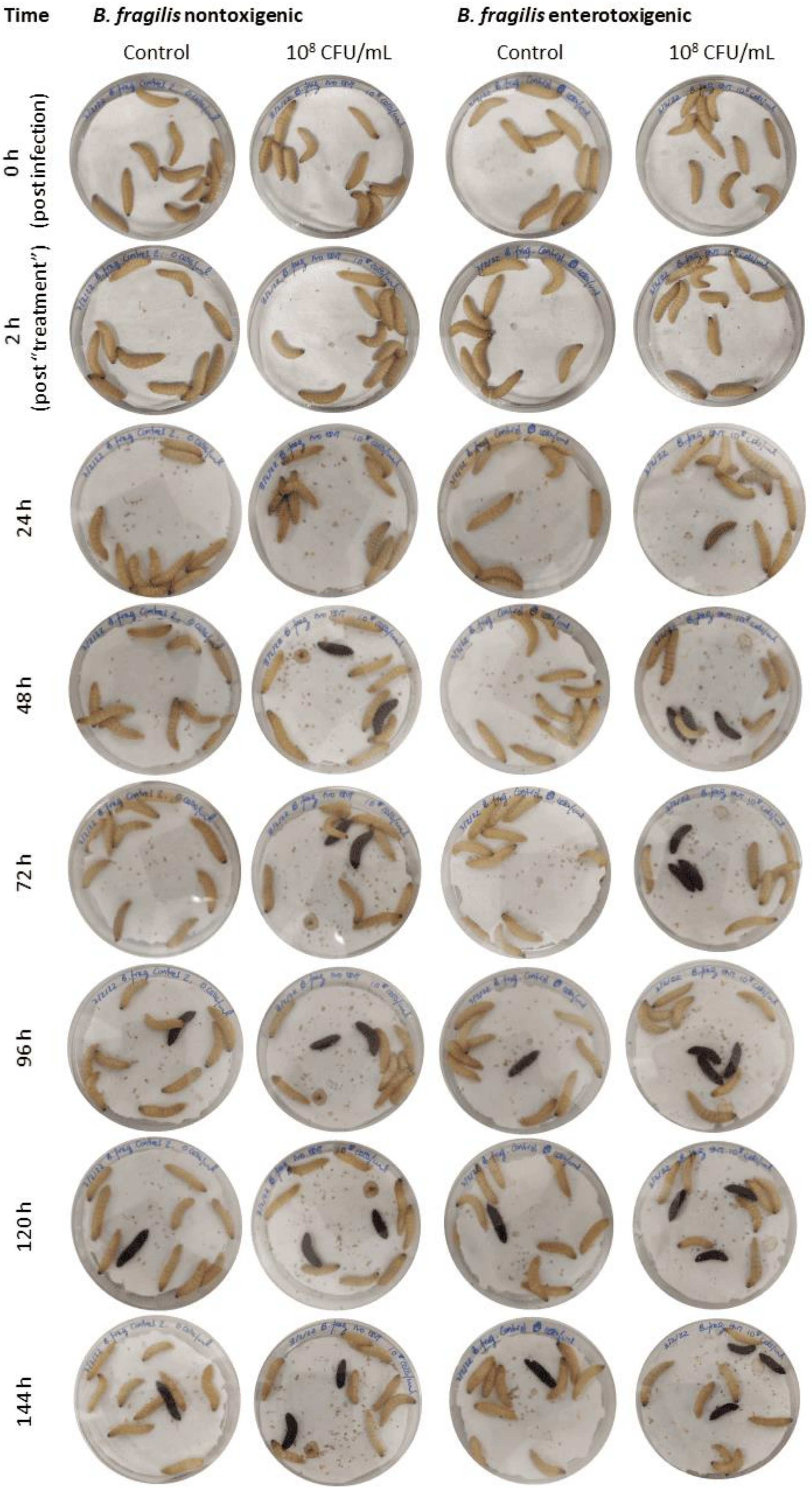

Time
B. fragilis nontoxigenic
B. fragilis enterotoxigenic
Control
10⁸ CFU/mL
Control
10⁸ CFU/mL
0 h (post infection)
2 h (post "treatment")
24 h
48 h
72 h
96 h
120 h
144 h

Figure S9. 24 h monitoring for determining the CFU/mL of *B. fragilis* necessary to produce a clear and rapid infection in *G. mellonella* larvae.

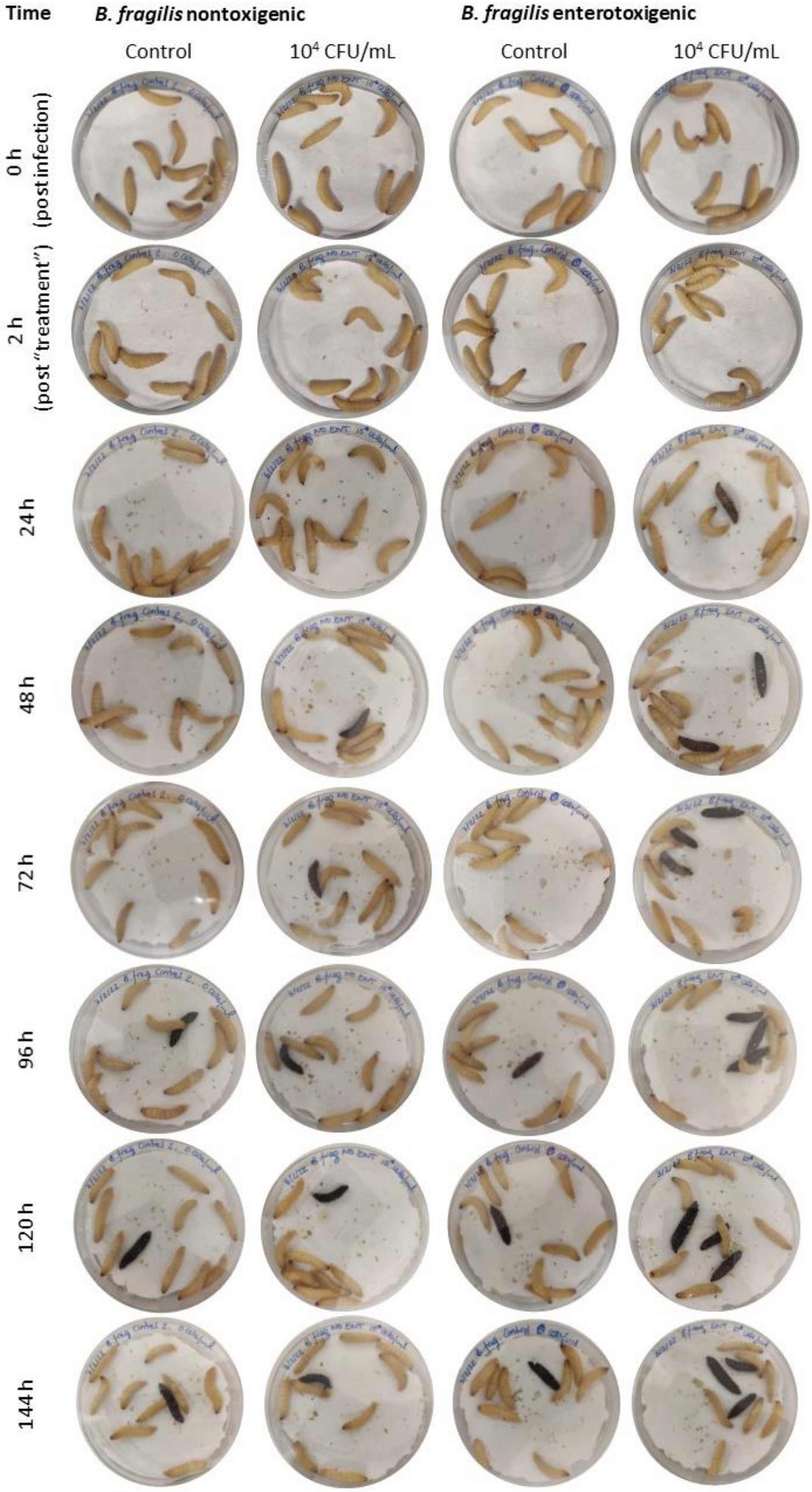

Time
B. fragilis nontoxigenic
B. fragilis enterotoxigenic
Control
10⁴ CFU/mL
Control
10⁴ CFU/mL
0 h (post infection)
2 h (post "treatment")
24 h
48 h
72 h
96 h
120 h
144 h

Figure S10. 24 h monitoring for determining the antibacterial effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae infected with both strains of *B. fragilis*.

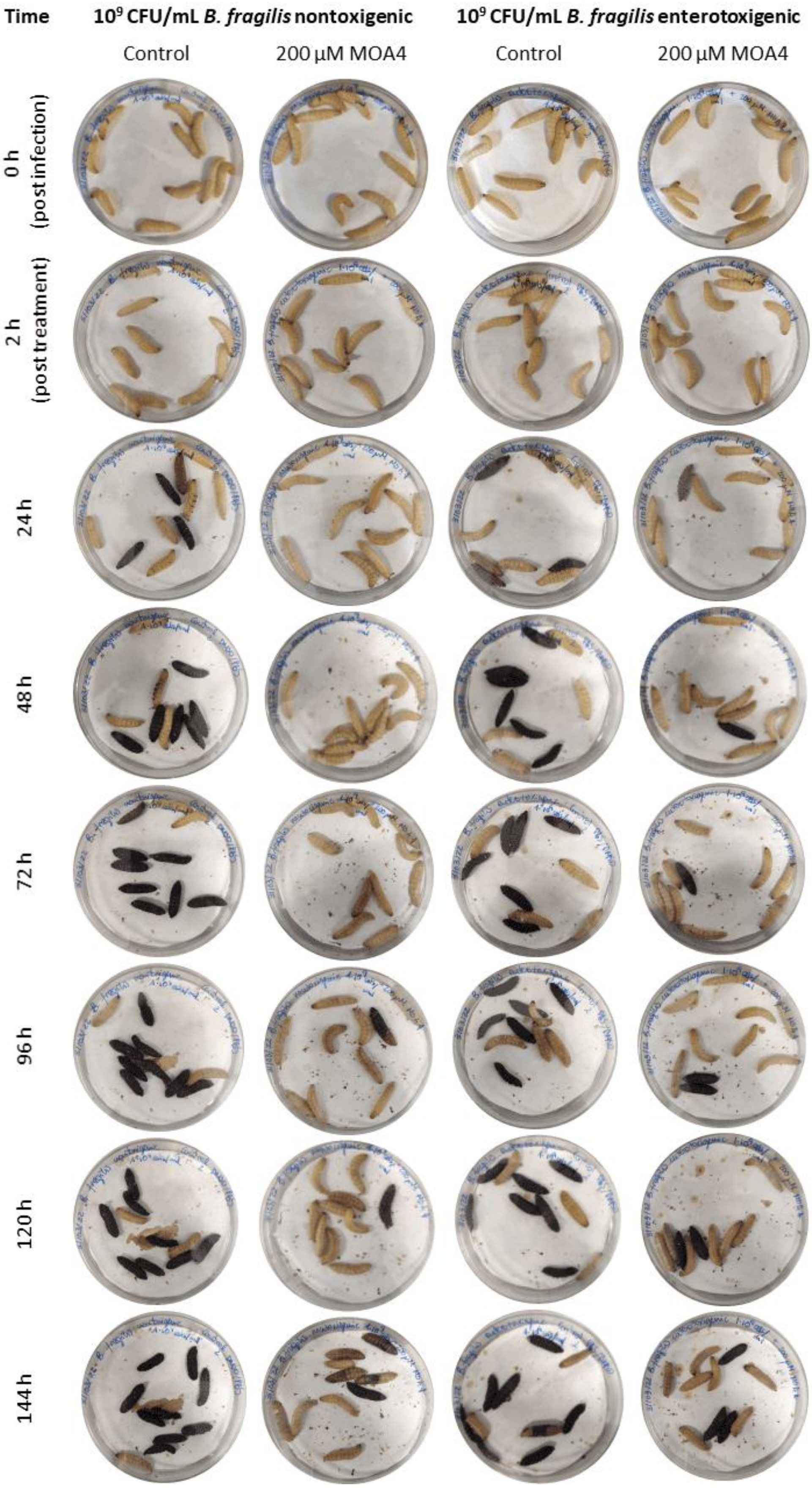

Time
10⁹ CFU/mL B. fragilis nontoxigenic
10⁹ CFU/mL B. fragilis enterotoxigenic
Control
200 µM MOA4
Control
200 µM MOA4
0 h (post infection)
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S11. 24 h monitoring for determining the antibacterial effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae infected with both strains of *B. fragilis*.

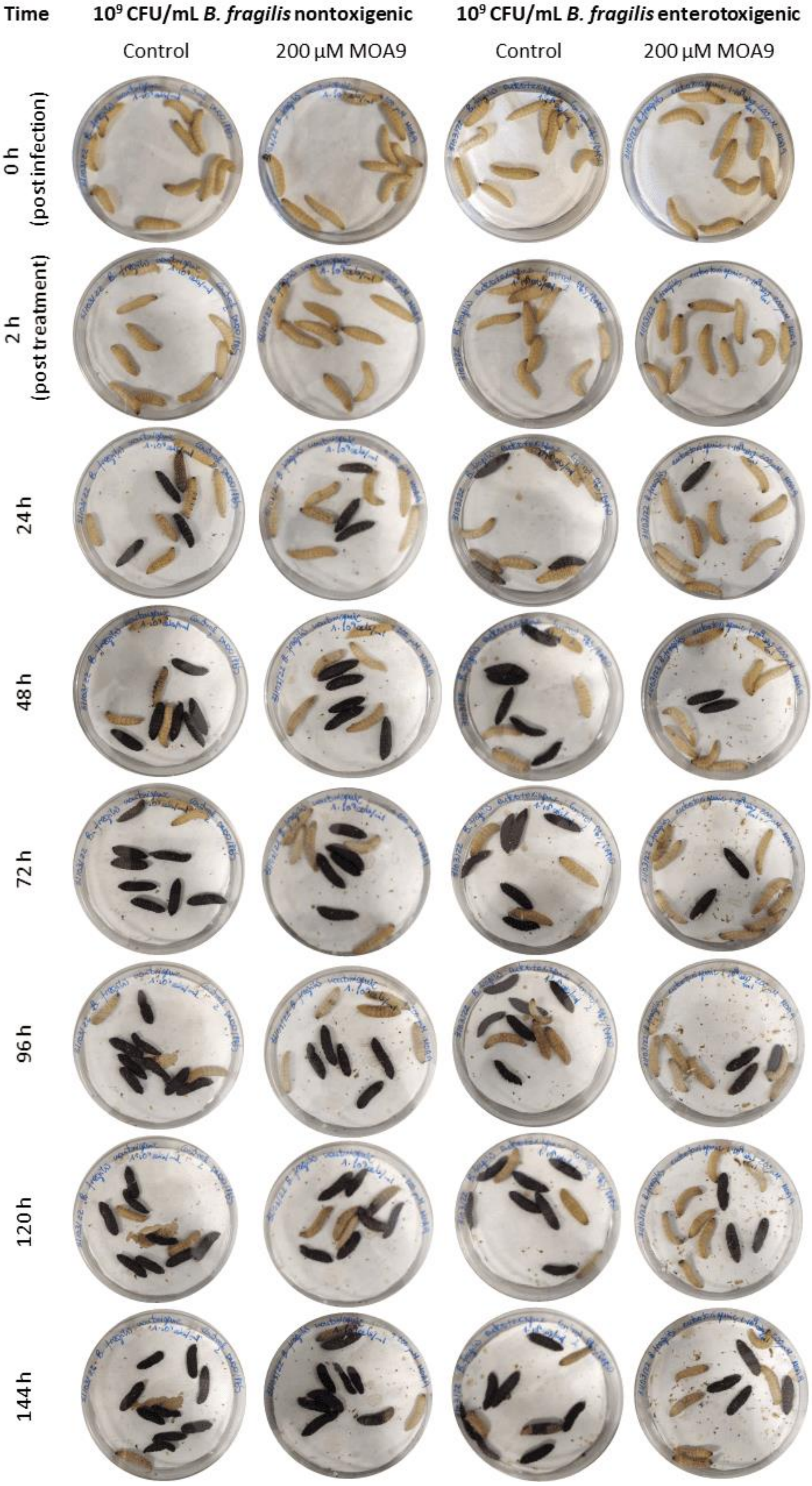

Time
10⁹ CFU/mL B. fragilis nontoxigenic
10⁹ CFU/mL B. fragilis enterotoxigenic
Control
200 µM MOA9
Control
200 µM MOA9
0 h (post infection)
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S12. 24 h monitoring for determining the antibacterial effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae infected with both strains of *B. fragilis*.

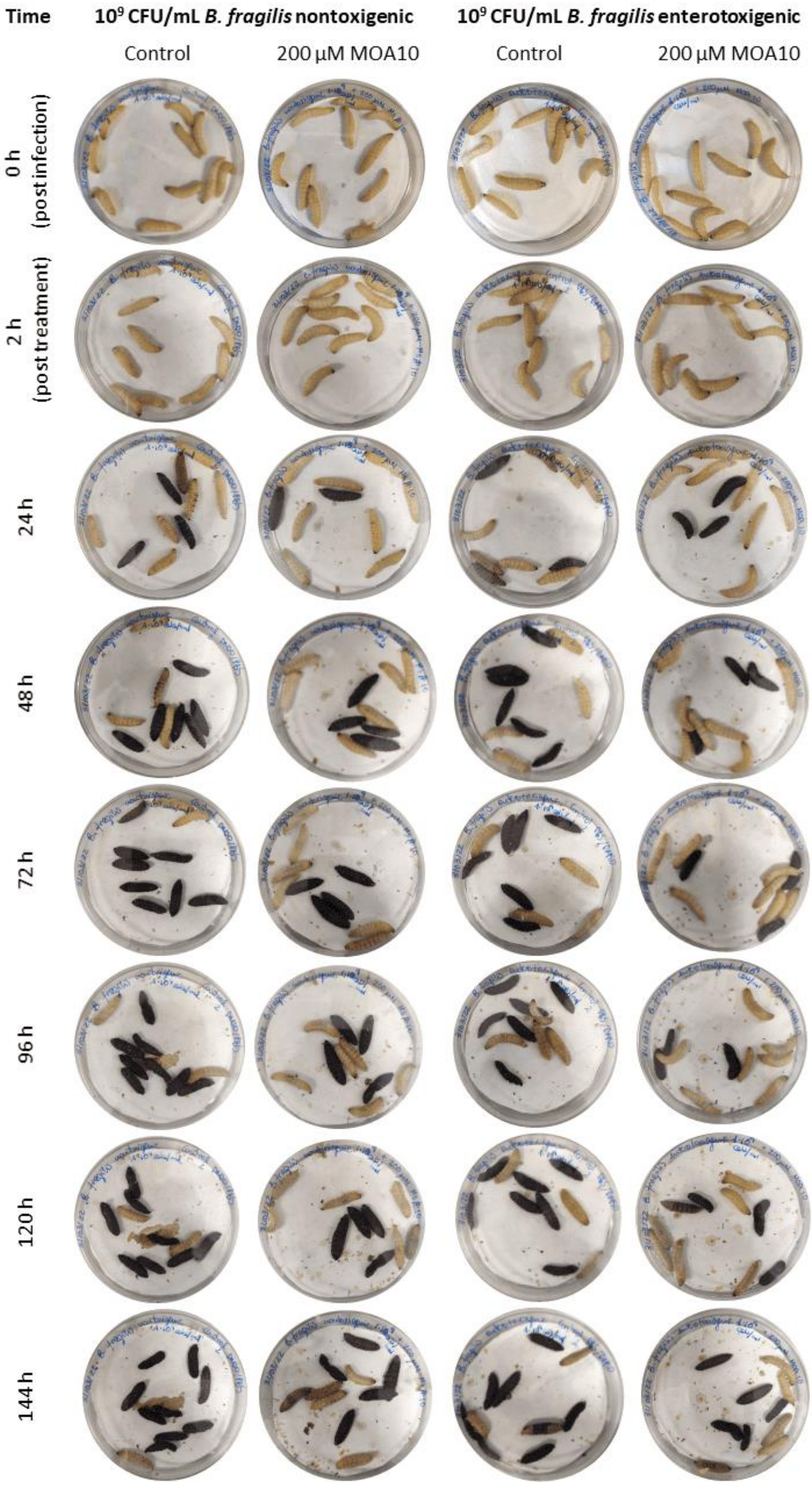
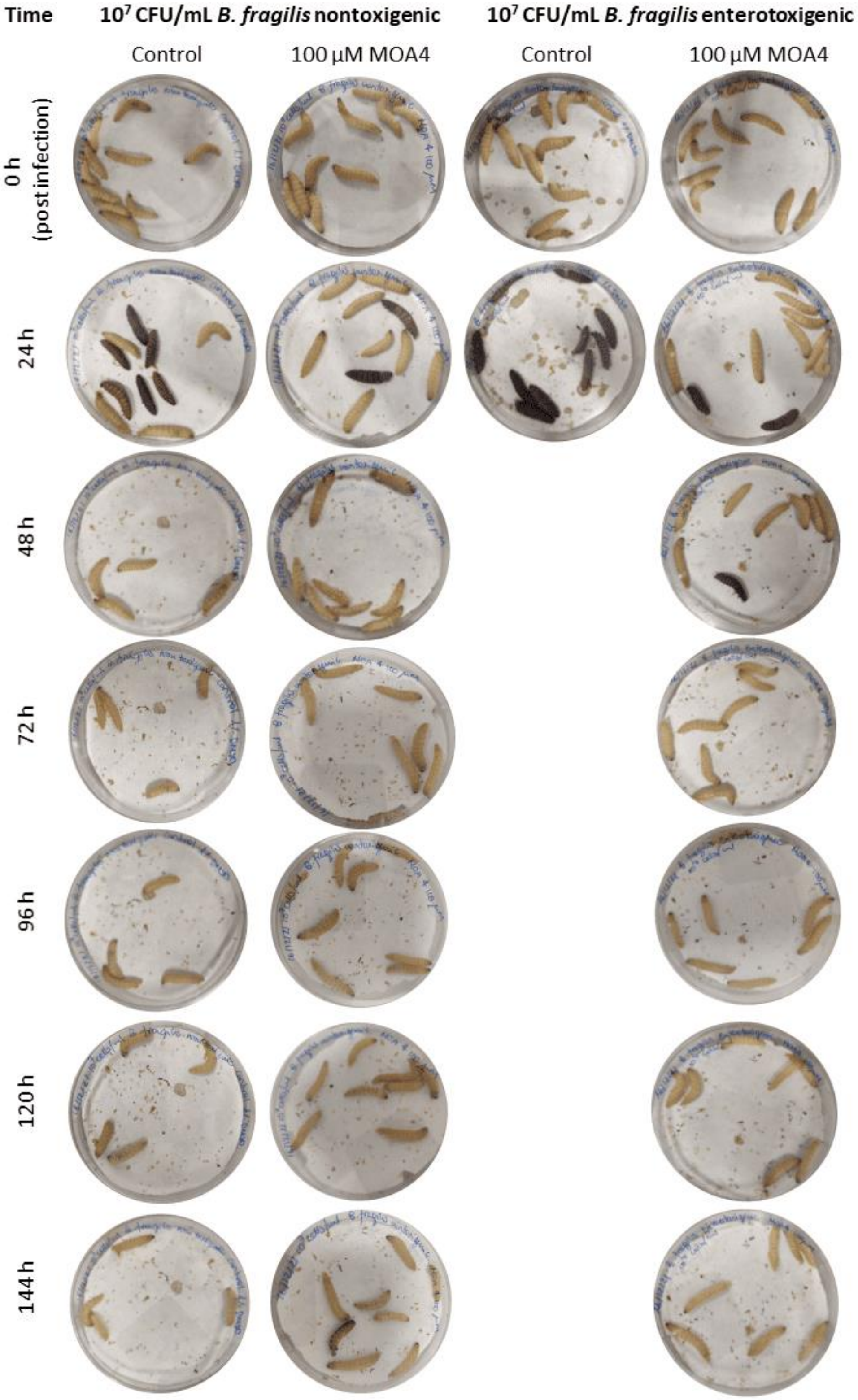

Time
10^9 CFU/mL B. fragilis nontoxigenic
10^9 CFU/mL B. fragilis enterotoxigenic
Control
200 µM MOA10
Control
200 µM MOA10
0 h (post infection)
2 h (post treatment)
24 h
48 h
72 h
96 h
120 h
144 h

Figure S13. 24 h monitoring for determining the antibacterial effect of MOA4, MOA9 and MOA10 compounds in *G. mellonella* larvae infected with both strains of *B. fragilis*.


Time
10⁷ CFU/mL B. fragilis nontoxigenic
10⁷ CFU/mL B. fragilis enterotoxigenic
Control
100 µM MOA4
Control
100 µM MOA4
0 h
(post infection)
24 h
48 h
72 h
96 h
120 h
144 h