



Trabajo de Fin de Grado

Grado en Biotecnología

Departamento de Bioquímica y Biología Molecular y Celular

Evaluación computacional de la energética de proteínas:
la flavodoxina de *Anabaena* sp. PCC 7119

Computational evaluation of protein energetics:
flavodoxin from *Anabaena* sp. PCC 7119

Francho Nerín Fonz

Directores: Juan José Galano Frutos y Javier Sancho

Universidad de Zaragoza

Índice

Portada.....	i
Índice.....	ii
Resumen	1
Abstract.....	1
1. Introducción	2
2. Contexto	4
2.1. Flavodoxina	4
2.2. Simulación molecular.....	6
2.3. Estabilidad de proteínas.....	9
2.4. Formación de complejo ligando-proteína	11
3. Materiales y métodos	13
3.1. Estructuras del PDB.....	13
3.2. Estructuras del estado desnaturalizado	13
3.3. Tamaño mínimo de muestra del estado intermediario y desplegado	14
3.4. Parámetros y etapas de simulación.....	14
4. Resultados y discusión.....	15
4.1. Estabilidad de la apoflavodoxina.....	15
4.2. Unión del FMN.....	19
5. Conclusiones.....	21
Conclusions	22
Bibliografía	23

Resumen

La información energética de simulaciones de Dinámica Molecular (DM) ha sido previamente utilizada para el cálculo preciso *in silico* de la estabilidad proteica de dos proteínas modelo, simplemente simulando sus estados nativo y desplegado. En este trabajo, se utiliza este método testado para el cálculo de la estabilidad proteica de una proteína modelo con un paisaje de estabilidad más complejo, con un intermediario de desplegamiento: la apoflavodoxina de *Anabaena* sp. PCC 7119. Se han conseguido reproducir *in silico*, de forma considerablemente exacta (dentro del margen de error), los parámetros termodinámicos de la proteína determinados *in vitro*, aunque para los cálculos realizados se necesitan las temperaturas medias de desnaturalización, fácilmente medibles *in vitro* (método “semi-computacional”). Además, la similitud entre los datos *in silico* e *in vitro* obtenida contribuye a demostrar tanto la precisión del campo de fuerzas y modelo molecular del agua utilizados en la DM (CHARM22/CMAP Y TIP3P) como la correcta representación del apenas poblado estado intermediario de la apoflavodoxina por parte de un mutante diseñado (PDB ID: 2KQU) y también del estado desplegado de las proteínas por parte de los *ensembles* generados por ProtSA. Sin embargo, al intentar extender el método al cálculo de la energética de unión de un ligando a una proteína (el FMN a la apoflavodoxina), no se han logrado obtener unos buenos resultados. A pesar de esto, se espera que las herramientas *in silico* continúen aumentando su precisión en el futuro y también se va a completar la validación de este método utilizando más proteínas modelo, además de volviendo a intentar la simulación de la unión de ligando, aplicando las distintas correcciones propuestas. Con todo, este método constituye una forma precisa y asequible de calcular la estabilidad proteica *in silico*, con importantes implicaciones en investigación básica y el campo de las proteínas terapéuticas y comerciales. (Recuento de palabras: 300)

Abstract

The energy data that can be collected from Molecular Dynamics (MD) simulations has been previously used to accurately calculate the stability of two model proteins, by simply simulating both the native and the unfolded state. In this work, this tested method is used for the calculation of the stability of a model protein with a more complex stability landscape, in which an unfolding intermediate appears: the apoflavodoxin from *Anabaena* sp. PCC 7119. The protein's thermodynamic parameters have been successfully reproduced *in silico* in great agreement, within the margin of error, with those determined *in vitro*, although the mean temperatures of denaturation, easily measured *in vitro*, are still needed for the calculation of the stability (“semi-computational” method). Moreover, the remarkable correspondence between the *in silico* and *in vitro* data assist to demonstrate both the precision of the force field and water model used in the MD simulations (CHARMM22/CAMP and TIP3P) and the adequate representation of the poorly populated intermediate state of the apoflavodoxin by a designed mutant (PDB ID: 2KQU) and also of the unfolded state of proteins by the *ensembles* generated by ProtSA. However, extending the procedure to the calculation of the binding of a ligand to a protein (FMN to apoflavodoxin) has not yielded the expected results, as it requires a 10-fold increase in the accuracy of the calculation. Despite this, *in silico* tools are expected to continue improving in the future, while this method is going to be subjected to complete validation using more model proteins as well as attempting (again) the ligand binding calculation, introducing the proposed adjustments. All in all, this method constitutes a precise and easily attainable procedure to calculate the stability of proteins *in silico*, with significant implications for basic research and therapeutic and commercial proteins. (Word count: 292)

1. Introducción

En los últimos años, el gran desarrollo de los métodos de **simulación molecular** ha permitido que hayan dejado de ser una técnica complementaria y se hayan convertido en una disciplina científica por sí misma¹. De forma similar, también ha habido un gran interés en el campo de la **estabilidad de proteínas**. Y, aunque en realidad el estudio de la estabilidad proteica es investigación básica, el conocimiento y las herramientas que se desarrollan como consecuencia son continuamente aplicadas en la creación de nuevos productos biofarmacéuticos y proteínas comerciales, donde la estabilidad proteica es de gran importancia². Tradicionalmente, el estudio de la estabilidad proteica se realiza mediante distintas técnicas *in vitro*, que miden (de forma directa o indirecta) algunas propiedades termodinámicas, como la variación de entalpía molar a presión constante (ΔH), la temperatura media de desnaturalización (T_m) o el calor específico molar a presión constante (C_p), con las que se calcula la estabilidad como la variación energía libre de Gibbs (ΔG), a través de la cual se relacionan^{2,3}.

En la evaluación *in vitro* de la estabilidad se suele medir la variación de los parámetros termodinámicos de la reacción química de desplegamiento (y/o plegamiento, si el proceso es reversible)^{1,2}. En esta reacción, se dan fenómenos como la formación y destrucción de interacciones no covalentes internas de la proteína y también de la proteína con el agua que la rodea¹. De forma análoga, todas estas interacciones (junto con muchos otros aspectos cuantitativos que conforman la energética de sistemas biomoleculares) están recogidos y modelados en los campos de fuerza que se utilizan en métodos *in silico*, como la Dinámica Molecular (DM), con los que potencialmente se pueden reproducir y cuantificar las energías computacionalmente⁴. Además, paralelamente al desarrollo de la capacidad computacional (que permite analizar problemas cada vez más complejos con estos métodos), también ha aumentado enormemente la precisión de los campos de fuerza y los modelos con los que se describe el agua, consiguiendo tanto una reproducción cada vez más fiel de los fenómenos biológicos como una cuantificación exacta de las energías en juego en los sistemas⁵.

Así, para evaluar el nivel de desarrollo de la Dinámica Molecular y la precisión de distintos campos de fuerza y modelos moleculares del agua, en **Galano-Frutos et al., 2019**⁶ utilizaron simulaciones cortas de DM de estructuras del estado nativo y desplegado de proteínas modelo para extraer de ellas datos energéticos con los que calcular la estabilidad. El grado de precisión de los modelos utilizados determina la exactitud de los valores obtenidos y, por lo tanto, también determina la similitud entre la ΔG conocida de experimentos *in vitro* y la calculada con este método, prácticamente por completo *in silico* (utiliza datos de T_m de técnicas sencillas *in vitro*, por lo que es un método “semi-computacional”).

En este **trabajo (Figura 1 A)**, se va a utilizar el mismo procedimiento aplicándolo a la apoflavodoxina de *Anabaena* sp. (syn. *Nostoc* sp.) cepa PCC 7119, una proteína modelo de desplegamiento en 3 estados termodinámicos. De esta manera, se obtendrán nuevas evidencias de la utilidad y fiabilidad de este método de cálculo de la estabilidad proteica *in silico* (haciéndolo además con una proteína con un mecanismo de desplegamiento más complejo), así como como de la precisión de los campos de fuerza y modelizaciones del agua que permiten hacerlo. La precisión en este caso es crítica, ya que los valores que se desean calcular tienen unas magnitudes mucho menores que los datos que se obtienen directamente de DM (que son los que se utilizan para los cálculos) y a su vez son menores que los de las proteínas modelo anteriores. Además, para ir más allá y demostrar la robustez del método, se va a extender su aplicación al cálculo de la estabilidad de unión de un ligando, en este caso el cofactor de la proteína, el flavín mononucleótido (FMN) (**Figura 1 B**).

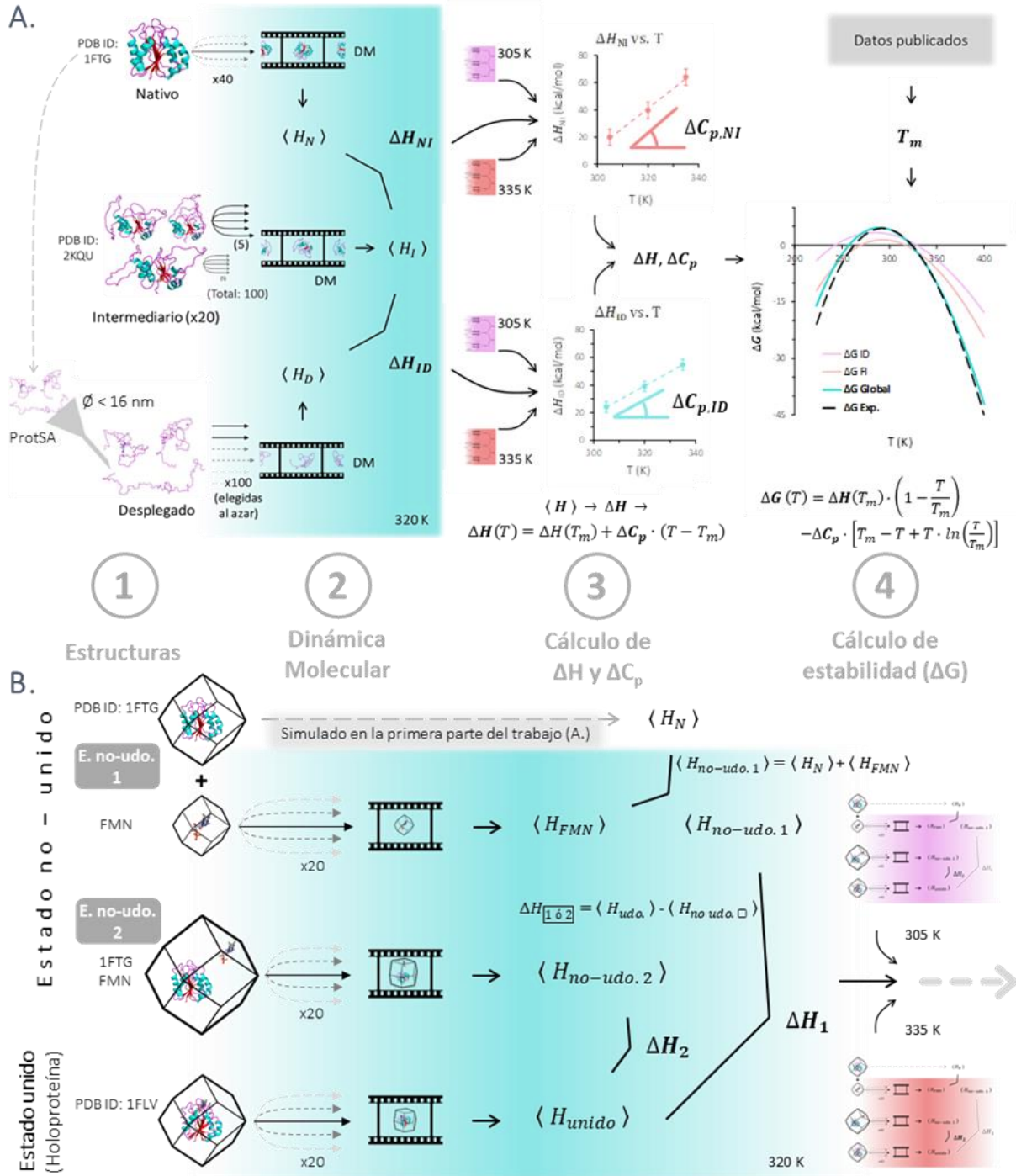


Figura 1. Flujo de trabajo utilizado para los cálculos de la estabilidad conformacional (A.) y de unión del FMN (B.) de la flavodoxina de *Anabaena* sp. **A.** 1. Estructuras del estado nativo (PDB ID: 1FTG⁷), intermediario (PDB ID: 2KQU⁸, 3 estructuras mostradas) y desplegado (obtenidas de ProtSA⁹ a partir del PDB del estado nativo y filtradas con un diámetro máximo de 16 nm). 2. Simulaciones de Dinámica Molecular (DM) de las que se extraen promedios de las entalpías absolutas ($\langle H \rangle$) de cada estado y a cada una de las tres temperaturas (305, 320 y 335 K). Para promediar, del estado nativo se realizan 40 réplicas a cada temperatura, del intermediario 5 réplicas de cada estructura a cada temperatura (100 totales a cada temperatura) y del desplegado 100 réplicas a cada temperatura a partir de 100 estructuras elegidas al azar del conjunto de estructuras filtradas. 3. Con las entalpías absolutas de cada estado y a cada temperatura ($\langle H \rangle$) se obtienen diferencias de entalpía (ΔH) a cada temperatura. La dependencia de (ΔH) con la temperatura permite obtener las diferencias de calor específico (ΔC_p) y las diferencias de entalpía a las temperaturas medias de desnaturalización de las dos transiciones (T_m), las cuáles se extraen de datos determinados in vitro publicados. 4. La estabilidad proteica se calcula con la ecuación de Gibbs-Helmholtz utilizando los datos calculados in silico y las temperaturas experimentales ya publicadas (método “semi-computacional”; gráfica morada) y se puede comparar con la estabilidad determinada in vitro (gráfica negra). **B.** 1. Estructuras utilizadas para la simulación del

estado unido (PDB ID: 1FLV¹⁰) y el estado no-unido. Para el estado no-unido, en la representación 1 se simula el FMN de forma independiente y luego las energías extraídas se suman a las de la apoproteína nativa (simulada en la primera parte del trabajo) mientras que en la representación 2 se simulan la apoproteína nativa (PDB ID: 1FTG⁷) y el FMN en el mismo sistema. **2.** También se realizan simulaciones de DM extrayendo entalpías absolutas a tres temperaturas, realizando 20 réplicas del FMN sólo, 20 réplicas del sistema conjunto apoproteína-FMN y otras 20 réplicas de la holoproteína. **3. y 4.** Se utilizan las mismas ecuaciones y relaciones termodinámicas que en la primera parte del trabajo para calcular la estabilidad de la unión y poder compararla con datos determinados por experimentos *in vitro*.

2. Contexto

2.1. Flavodoxina

Las **flavodoxinas** son proteínas pequeñas de un único dominio de plegamiento involucradas en la transferencia de electrones, como en la fotosíntesis, y que contienen una molécula de FMN no covalentemente pero fuertemente unida. Se descubrieron en 1960 en cianobacterias y bacterias de la clase *Clostridia*, siendo prácticamente exclusivas de bacterias (también se encuentran en algunas algas verdes y rojas). En ellas, sustituyen a la ferredoxina en ambientes pobres en hierro, pero también están involucradas en una gran variedad de reacciones, de manera que en algunos organismos son esenciales y se expresan de forma constitutiva. Actualmente, se ha intensificado su investigación como proteína modelo de estudios de estabilidad y plegamiento proteicos, unión reversible de ligando y en algunos casos como diana de compuestos antibacterianos¹¹.

La flavodoxina de *Anabaena sp.* pertenece al grupo de las flavodoxinas de cadena larga, que tienen un loop adicional respecto a las de cadena corta, cuya función es hasta ahora desconocida. También, se ha propuesto la pertenencia de los dos tipos de flavodoxina a dos distintos grupos evolutivos diferenciados, de los que habría aparecido en primer lugar el de las de cadena larga¹¹.

Es una proteína sencilla monodominio con la estructura típica α/β y con tres capas, siendo la central una lámina β paralela de cinco hebras (**Figura 2 A**). La única diferencia estructural entre la holo y la apoproteína es uno de los bucles que interacciona con el FMN, que en su ausencia se desplaza para cerrar el espacio que queda y que al cristalizar suele retener un anión del tampón que mimetiza al fosfato del FMN¹¹.

El plegamiento y unión a FMN son procesos secuenciales y éste último sólo se da con la proteína plegada, por lo que la presencia de FMN no acelera la cinética de plegamiento. Además, en el plegamiento se acumula de forma transitoria un intermediario cinético fuera de la ruta principal que reduce la velocidad de la reacción, mientras que en el plegamiento-desplegamiento térmico aparece un intermediario termodinámico distinto al cinético¹¹.

La unión del **FMN** se produce por interacciones específicas de apilamiento de orbitales π del anillo de isoaloxazina con un triptófano y una tirosina conservados (**Figura 2 B**), mientras que el fosfato interacciona de forma no específica por enlaces de hidrógeno con las cadenas laterales de varios residuos de treonina y varios hidrógenos de enlaces peptídicos (**Figura 2 C**) y la parte del ribitilo no colabora significativamente. Sin embargo, la manera precisa por la que se produce no está del todo determinada. La unión se produce en la parte de la proteína que se despliega en el intermediario y

aumenta su estabilidad conformacional, por lo que en el desplegamiento térmico la holoproteína sigue un mecanismo simple de dos pasos¹¹.

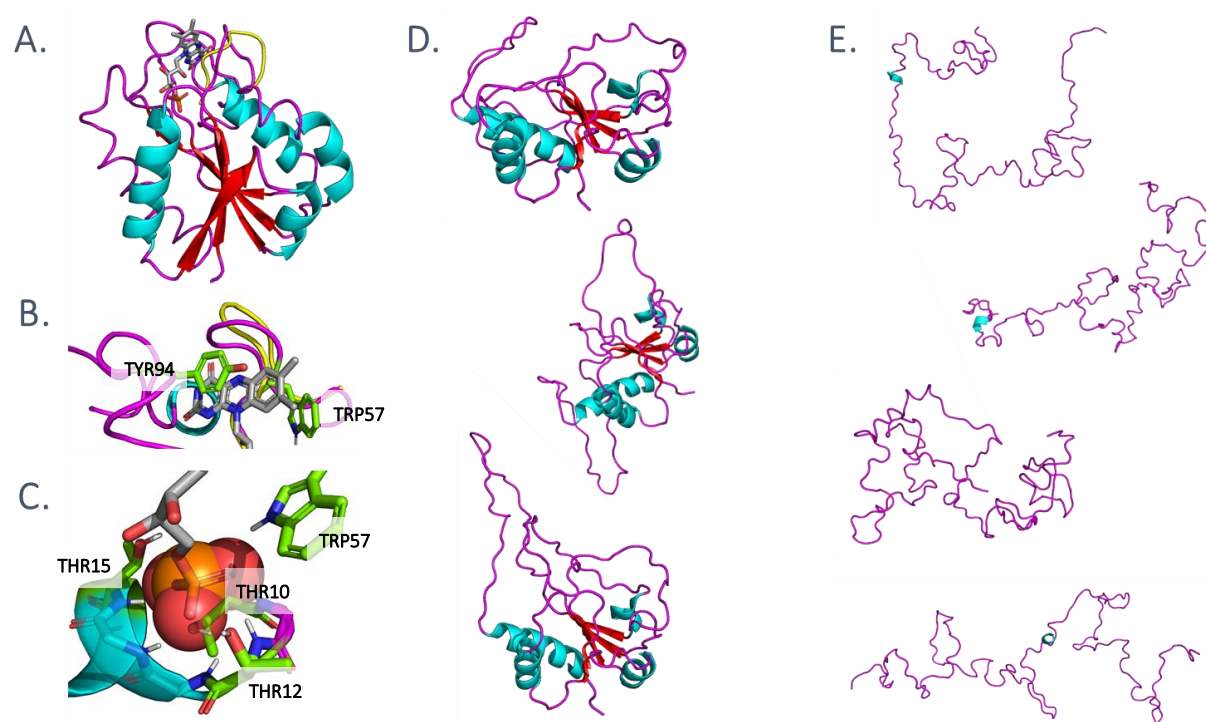


Figura 2. Estructuras de la flavodoxina de *Anabaena sp.* utilizadas en las simulaciones, todas coloreadas por estructura secundaria: hebras β rojas, hélices α azules y bucles fucsias. **A.** Estructura de la holoproteína (PDB ID: 1FLV¹⁰), con el cofactor FMN en gris, coloreado por elemento químico. En amarillo está representada la conformación correspondiente al bucle de unión a FMN en la apoproteína (PDB ID: 1FTG⁷), que en ella está desplazado para cerrar el hueco que deja la ausencia del FMN y que constituye la diferencia más significativa entre las dos estructuras. **B.** Detalle de la interacción del triple anillo de isoaloxazina del FMN con los residuos TRP57 y TYR94. **C.** Detalle de la interacción del fosfato del FMN con las cadenas laterales de los residuos de treonina 10, 12 y 15, con los hidrógenos de varios enlaces peptídicos y también con el amino secundario de la cadena lateral del TRP57. **D.** Representación de 3 de las 20 estructuras del estado intermedio (PDB ID: 2KQU⁸). **E.** Representación de 4 estructuras del ensemble de estructuras del estado desplegado obtenido en ProtSA⁹.

Al igual que muchos intermediarios de proteínas que se despliegan térmicamente en tres estados, el **intermediario** de la flavodoxina de *Anabaena sp.* tiene una energía libre cercana a la del estado nativo. Por lo tanto, durante el proceso, nunca adoptan la estructura más de un 65% de las moléculas, por lo que no se puede estudiar directamente con Resonancia Magnética Nuclear (RMN) en solución. Además, la conformación que adopta a bajo pH o el estado de transición del proceso de plegado se ha determinado que no se parecen a él. Por otro lado, las superficies internas de las proteínas son comúnmente apolares, lo que favorece el empaquetamiento hidrofóbico, pero la superficie de la flavodoxina entre el dominio que se mantiene plegado y el que se desordena primero durante el desplegamiento térmico es anormalmente polar, lo que disminuye su ordenamiento por la incapacidad de formar núcleos hidrofóbicos⁸.

Combinando datos de DM y espectroscopía, se determinó que el intermediario aparece enseguida en el proceso de desplegamiento y que se produce por una deformación estructural de los bucles y luego la transferencia de esta perturbación a los elementos de estructura secundaria que los rodean. Esto sugiere (junto con la uniformidad de velocidad de desplegamiento) que siguen este único mecanismo a nivel microscópico. El intermediario no aparece en la desnaturalización química con urea, ya que la urea

ataca directamente al núcleo hidrofóbico, mientras que éste se mantiene en los primeros pasos de la desnaturalización térmica¹².

Para estabilizar la conformación intermediaria y que a unas condiciones determinadas la mayor parte de las moléculas la adopten y se pueda determinar su estructura por RMN en solución, se generó un mutante de la proteína cambiando un aminoácido de la interfase entre los dominios plegado y desplegado del intermediario. Se sustituyó una fenilalanina de un núcleo hidrofóbico de la interfase por una asparagina, provocando una gran desestabilización del estado nativo pero que no afectaba al estado intermediario, en el que la posición pertenece a la parte desplegada y probablemente interacciona favorablemente con el solvente. En condiciones suaves, la proteína resultante adquiere una estructura parecida a la del intermediario, sin poblar significativamente el estado desplegado (según datos de dicróismo circular de UV cercano y lejano), aunque se generan nuevas interacciones que aumentan levemente la organización. El conjunto de estructuras obtenidas por RMN (**Figura 2 D**) también se refinó con datos de Dispersión de Rayos X de Ángulo Reducido (*Small-Angle X-ray Scattering*, SAXS)⁸.

2.2. Simulación molecular

La **química computacional** se ha vuelto cada vez más versátil. Empezando por la química cuántica, con la que se describen los orbitales moleculares y los electrones que los ocupan, permitiendo calcular cualquier propiedad física o química que dependa de la distribución de electrones, hasta las simulaciones de dinámica molecular de grano grueso, en las que se utilizan “esferas” que agrupan y describen grupos de átomos y sus interacciones según las leyes de la mecánica newtoniana, permitiendo simplificar la descripción de los sistemas y obtener información de sistemas biológicos más complejos, incluso a nivel celular. El tamaño máximo de sistema analizable con química cuántica no supera los cientos de átomos, mientras que con métodos basados en la mecánica molecular, en los que la energía se describe de forma mucho más simple, se llega a los cientos de miles y, dentro de ellos, con los de grano grueso a varios millones de átomos⁴.

Para la mayoría de problemas, la mejor alternativa es una técnica experimental, pero debido a los avances recientes puede haber casos en los que las simulaciones aportan mayor detalle o son más eficientes que plantear un nuevo experimento. Además, aunque tradicionalmente las simulaciones se han utilizado para comprobar si los modelos teóricos en los que se basan pueden reproducir observaciones experimentales o aportar un mayor nivel de detalle, últimamente se utilizan por sí solas para realizar predicciones y posteriormente realizar la confirmación de forma experimental¹.

En las simulaciones convencionales, las moléculas se describen con **campos de fuerza**, que agrupan tanto a las ecuaciones de mecánica molecular clásica utilizadas para describir los movimientos de los átomos, sus interacciones y energía como a los parámetros cuantitativos contenidos en éstas, que caracterizan el coste energético en relación al valor en equilibrio para que se produzca una variación de la interacción descrita. Las interacciones se dividen en las enlazantes o covalentes y las no enlazantes o no covalentes. Las covalentes incluyen la vibración de enlaces y ángulos de enlace y las torsiones entre 4 átomos unidos de forma consecutiva (*dihedrals*) o no consecutiva (*improper dihedrals*) y las no covalentes las de van der Waals (según el modelo de Lennard-Jones) y las electrostáticas (Coulomb)¹. La simplicidad de la forma de cálculo del funcional de energía supone que, por un lado, se puedan hacer cálculos rápidos y fáciles, pero que, por otro, la exactitud de los cálculos dependa de los parámetros

utilizados incluidos en los términos de energía. Éstos pueden obtenerse de cálculos de química cuántica *ab initio*, del ajuste de datos experimentales o de métodos semi-empíricos⁴.

Los campos de fuerzas pueden variar en las formas matemáticas utilizadas para describir los distintos tipos de energías e interacciones y, obviamente, en el valor de sus parámetros asociados. El campo de fuerzas utilizado aquí es CHARMM, un campo de fuerzas *all-atom* y aditivo (suma las energías parametrizadas por separado) que surgió en los 80 y que ha alcanzado un grado significativo de compleción, incluyendo, además de proteínas, ácidos nucleicos, lípidos y carbohidratos y también la posibilidad de parametrizar moléculas orgánicas, como ligandos y fármacos¹.

Los **métodos de simulación** más utilizados para obtener *ensembles* son Monte Carlo (MC, estocástico/aleatorio) y Dinámica Molecular (DM, determinístico) que integran las ecuaciones de Newton de que relacionan fuerza y aceleración para calcular los movimientos y posiciones de los átomos y la energía del sistema. Por otro lado, con el método más simple de minimización energética, por ejemplo el de descenso de máxima pendiente (*steepest descent*), los átomos simplemente se mueven de forma que se minimiza la energía del sistema, y es un paso previo necesario para obtener sistemas equilibrados previamente a la utilización de los métodos de simulación¹.

En una simulación de **DM**, el tiempo se discretiza y la posición tras un pequeño tiempo finito sobre el que actúan unas determinadas fuerzas se puede calcular de forma relativamente sencilla. Para el cálculo, se necesitan conocer las posiciones y velocidades de partida de los átomos del sistema, siendo las posiciones normalmente las de una estructura aportada (generalmente un archivo pdb) y las velocidades las que se asignan de forma aleatoria, según la distribución de Maxwell-Boltzmann en función de la temperatura (y por lo tanto de la energía cinética). Con ellas, la fuerza ejercida sobre cada átomo es calculada y de ella se obtiene la dirección, sentido y magnitud del movimiento, con lo que se obtiene la nueva posición de cada átomo tras el pequeño paso de tiempo definido en el que la fuerza se supone constante, y se repite el proceso el número de pasos deseado, obteniendo información a lo largo del tiempo y por lo tanto también reproduciendo fenómenos cinéticos y dinámicos⁴.

Para los sistemas con más de tres átomos interaccionantes y sin restricción de los grados de libertad no hay una solución analítica de las ecuaciones de movimiento para los pasos de tiempo discretos, por lo que resuelven con algoritmos de integración numérica o **integradores**, como Leapfrog, el utilizado en este trabajo. Deben introducir el mínimo error posible, tanto por las aproximaciones en las que se basan como por la pérdida de precisión a lo largo de los cálculos, y ser computacionalmente eficientes. En la eficiencia también influye la magnitud del paso de tiempo elegido, que cuanto mayor sea mayor tiempo permitirá simular con una capacidad computacional dada y viceversa. Sin embargo, está limitado por el movimiento más rápido del sistema (la vibración de enlace entre un carbono y un hidrógeno), aunque se pueden utilizar estrategias, para poder aumentar el paso de tiempo, como la fijación de las distancias de enlace con algoritmos como LINCS, utilizado en este trabajo. Aun solventando este problema, el paso de tiempo tampoco puede ser excesivamente grande, ya que algunos átomos del sistema podrían desplazarse a lo largo de distancias considerables y provocar la aparición de solapamientos en las nuevas posiciones tras la integración, con lo que aparecerían también energías muy altas y el sistema podría acabar “explotando”¹³. Además, al utilizar un campo de fuerzas de mecánica clásica, no se permite la rotura ni formación de enlaces y también se fija el pH¹³, asignando los estados de protonación de los grupos relevantes en función de su entorno según lo calculado por PROPKA, utilizando el servidor PDB2PQR¹⁴.

La simulación comienza tras la **minimización** con la etapa de **calentamiento**, concatenando varios pasos de simulación con los que se llega hasta la temperatura deseada, aumentándola progresivamente hasta alcanzarla. Luego, se realiza el **equilibrado**, también realizando varios pasos para terminar de ajustar las condiciones (por ejemplo, energía, temperatura y presión) a las que se quiere realizar la **producción** propiamente dicha, para que ésta parta de un punto estable. El comienzo de la recogida de datos nunca debe comenzar tras un cambio en las condiciones de simulación, por lo que el último paso del equilibrado debe realizarse en las mismas condiciones en las que se realizará la producción o se deben descartar los primeros datos obtenidos en la producción. Tanto para el mantenimiento de las condiciones de simulación como para otros aspectos, como la creación del sistema, la descripción de sus átomos e interacciones o el cálculo de las energías y posiciones, se utilizan múltiples aproximaciones matemáticas y físicas y algoritmos de cálculo¹³.

A diferencia de los sistemas macroscópicos, idealmente diluidos con gran cantidad de moléculas de agua y con billones de moléculas de soluto que forman un *ensemble*, en las simulaciones sólo se estudia una molécula y además en una cantidad reducida de moléculas de agua. La simulación de una sola molécula en principio no es representativa de todo el *ensemble*, pero, si se simula durante suficiente tiempo, puede permitir aplicar la hipótesis de **ergodicidad** y obtener propiedades medias del *ensemble* como la media a lo largo del tiempo de simulación. Incluso con los avances recientes en rapidez y adaptación de los cálculos a entornos de computación distribuida, los tiempos de simulación realizables con DM siguen siendo significativamente más cortos que los tiempos típicos de muchos procesos biomoleculares y técnicas experimentales. Para obtener mayor cantidad de información con los recursos actuales y/u obtener una mejor representación de los *ensembles*, se han desarrollado varios métodos de muestreo mejorado (*enhanced sampling*), como *Metadynamics* o *Replica Exchange*¹.

Por otro lado, la utilización de una cantidad reducida de agua provoca la aparición de artefactos debido a la existencia de límites del sistema, lo que se corrige utilizando **Condiciones Periódicas de Contorno** (*Periodic Boundary Conditions*, PBC). Se sitúan réplicas del sistema de forma periódica en las tres direcciones del espacio para representar un sistema pseudo-infinito, eliminando el vacío, y que así las moléculas que salgan de él por un límite entren por el opuesto correspondiente¹.

Aun así, el sistema debe ser suficientemente grande para que las distintas moléculas no interactúen con ellas mismas, sobre todo por **interacciones no enlazantes o no covalentes** y entre éstas las electrostáticas, que se dan a larga distancia. Las interacciones Lennard-Jones decaen rápidamente con la distancia y se pueden establecer valores límite a partir de los que se dejan de evaluar, pero la introducción de estos tipos de valores límite para las coulombicas produce errores y por ello se utiliza el método de malla de partículas de Ewald (*Particle-Mesh-Ewald*, PME). Con PME, las interacciones electrostáticas se dividen entre una de corto alcance y, a partir de un valor límite, una de largo alcance, que se trata asignando cargas a una representación del sistema dividido en una malla y que se resuelve en el espacio recíproco mediante transformadas de Fourier¹. También se suelen despreciar las interacciones no enlazantes o no covalentes entre átomos vecinos (*exclusions*), para que sólo se vean afectados por sus enlaces y ángulos de enlace y torsiones, además de las interacciones con el resto¹³.

Para lograr una simulación a temperatura constante, se utilizan **termostatos** que modifican la energía cinética para que sea proporcional a la temperatura deseada (según el teorema de equipartición que los relaciona). En este trabajo, en el calentamiento se utiliza el termostato Berendsen, que modifica las velocidades de los átomos de forma progresiva y suavizada para que el sistema esté exactamente a la temperatura fijada, y se prosigue utilizando el Bussi-Donadio-Parrinello, que produce menos artefactos

y afecta menos a procesos dinámicos y también modifica las velocidades, en este caso para lograr una energía cinética elegida de forma estocástica/al azar en función de la distribución de energía cinética del *ensemble* canónico (a temperatura constante)¹³.

Para mantener la presión y permitir la variación de volumen, también se utilizan **barostatos** junto con los termostatos, simulando el *ensemble* NPT (con número de átomos, presión y temperatura constantes). El mantenimiento de la presión se comienza a aplicar en un paso del equilibrado con el barostato Berendsen, que, de forma similar al termostato del mismo nombre, varía el volumen de forma progresiva y suavizada para mantener la presión. Posteriormente, se cambia al Andersen (o Parrinello-Rahman isotrópico), que añade un término adicional a las ecuaciones de movimiento de los campos de fuerza con el que simula la acción de un pistón que actúa de forma uniforme en las tres direcciones del espacio (isotrópico)¹³.

2.3. Estabilidad de proteínas

Las proteínas son las biomoléculas que llevan a cabo la mayoría de las tareas en los seres vivos, y de ello es responsable la **estructura tridimensional** específica que adoptan de entre la enorme cantidad de conformaciones posibles, incluso si sólo lo hacen en momentos puntuales, cómo al interactuar con otras moléculas en el caso de las proteínas intrínsecamente desordenadas². Ya que el polipéptido se podría plegar en muchas otras conformaciones, el estado nativo representa el mínimo de energía libre de Gibbs (ΔG) de todos los posibles estados¹⁵.

La estructura tridimensional concreta que adoptan las proteínas está determinada por las propiedades fisicoquímicas de su secuencia de aminoácidos y en su adopción intervienen fuerzas como los enlaces de hidrógeno, interacciones de van der Waals y electrostáticas, la adopción de ángulos de enlace peptídico preferentes, empaquetamiento hidrofóbico y reducción de la entropía. El plegamiento ocurre de manera muy **rápida**, por lo que el proceso no ocurre explorando todas las posibles conformaciones de forma aleatoria, lo que llevaría mucho tiempo. Las proteínas se pliegan tan rápido sin explorar todo su espacio conformacional porque éste tiene una forma de “embudo” (**Figura 3 A**) por el que, si la proteína adquiere conformaciones al azar, generalmente son cada vez más estables y con cada cambio conformacional se va acercando al mínimo energético, existiendo en cada nivel cada vez menos conformaciones posibles y por ello alcanzando el mínimo rápidamente (mientras que existen muchas estructuras desplegadas). Aún con esto, las características locales del paisaje conformacional, que son las que se investigan en simulaciones moleculares, pueden tener grandes barreras energéticas que impiden que éstas se puedan utilizar para observar el plegamiento⁵.

Sin embargo, no se ha establecido un **mecanismo de plegado general** o características generales del proceso que se puedan observar en varias proteínas. Parece que las proteínas se pliegan en unidades de estructura secundaria, aumentando su estabilidad al aumentar la cantidad de estructura secundaria que se va plegando durante el proceso, y que primero se desarrollan estructuras locales que luego crecen en otras estructuras más globales. Además, diferentes moléculas de una proteína pueden seguir caminos de plegado a través del paisaje energético diferentes a nivel microscópico con el mismo resultado, también pudiendo haber algunos caminos más poblados que otros (**Figura 3 A**)⁵.

Para el **estudio *in vitro*** de la estabilidad proteica, se utilizan distintas técnicas que miden, de forma directa o indirecta, distintas propiedades termodinámicas, como la Calorimetría Diferencial de Barrido

(*Differential Scanning Calorimetry*, DSC). Normalmente, estas mediciones se realizan a concentraciones crecientes de un agente desnaturante, bien de naturaleza química o un aumento de temperatura, y las variaciones de los parámetros termodinámicos que se registran se deben a la reacción de desplegamiento (o plegamiento)^{1,2}. El desplegamiento se trata de una reacción química de cambio de fase de primer orden (que por lo tanto tiene una barrera energética (**Figura 3 B**) y también una cinética del proceso), en la que se pasa de un estado nativo, estabilizado entálpicamente por interacciones internas de la proteína y de ésta con el solvente, al estado desplegado, estabilizado entrópicamente gracias al gran número de conformaciones que puede adoptar en él¹.

Para cada proteína existe una constante de equilibrio entre la fracción de moléculas plegadas y desplegadas que además es también la relación entre las constantes cinéticas de los procesos de plegamiento y desplegamiento. Un desplegamiento rápido puede provocar que se inactive mayor cantidad de una proteína en un mismo periodo de tiempo, respecto de una proteína con la misma constante de equilibrio pero con distinta cinética; mientras que uno lento proporciona una mayor estabilidad cinética, que suele ser la responsable de que no se formen agregados β -amiloideos pese a que sean más estables termodinámicamente que el estado nativo².

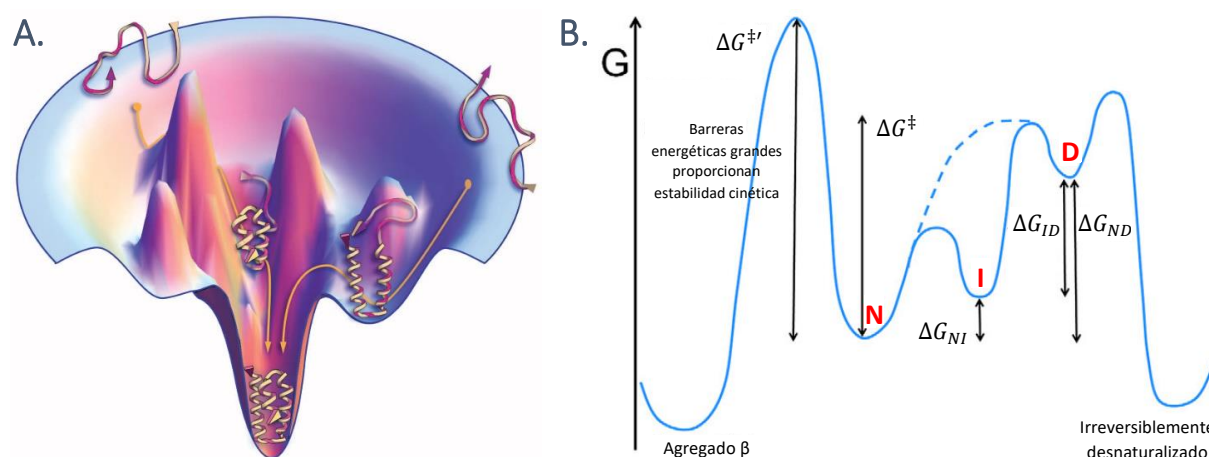


Figura 3. **A.** Representación tridimensional del paisaje conformacional de las proteínas en forma de “embudo” y ejemplo de dos posibles caminos microscópicos de plegamiento que convergen en el estado nativo. Tomado de Dill y MacCallum, 2012⁵. **B.** Ejemplo de una sección bidimensional del paisaje conformacional de una proteína con un mecanismo de desplegamiento en tres estados (o de dos estados, siguiendo la línea discontinua) representando también mínimos energéticos de agregados β y del estado desnaturizado irreversible. Adaptado (traducido) de Sancho, 2013².

Normalmente, en los equilibrios de proteínas pequeñas las únicas conformaciones que se pueblan de forma medible son la plegada o nativa (N) y desplegada (D), pero para muchas proteínas existen **intermediarios** (I) estables adicionales². Estas conformaciones intermedias tienen propiedades mixtas entre los estados nativo y desnaturizado y aparecen en condiciones que desestabilizan la conformación plegada pero no estabilizan del todo la desplegada, como en presencia de una alta concentración de solutos o como consecuencia de mutaciones. Estos intermediarios pueden formar parte de la ruta de desplegamiento, facilitando la reacción o como “trampas termodinámicas” que pueden llevar a malplegamiento o incluso agregación (**Figura 3 B**)¹². El número de intermediarios que aparecen y se detectan depende del paisaje energético y del agente desnaturante utilizado y no son fáciles de detectar ni tampoco de estudiar a nivel atómico, debido a su baja población en condiciones normales^{2,12}. El equilibrio resultante se puede dividir en equilibrios parciales para cada transición, con

diferentes constantes y por lo tanto diferencias de energía libre para cada uno de ellos. Para una proteína de 3 estados, los equilibrios y la energía libre de Gibbs resultante es:

$$N \leftrightarrow I \leftrightarrow D \quad ; \quad K_{NI} = \frac{\chi_I}{\chi_N} \quad ; \quad K_{ID} = \frac{\chi_D}{\chi_I} \quad ; \quad \Delta G_{\boxed{NI \text{ ó } ID}}(T) = -R \cdot T \cdot \ln K(T) = \Delta H - T \cdot \Delta S$$

Para este trabajo, se utiliza la ecuación integrada de Gibbs-Helmholtz:

$$\Delta G_{\boxed{NI \text{ ó } ID}}(T) = \Delta H(T_m) \cdot \left(1 - \frac{T}{T_m}\right) - \Delta C_p \cdot \left[T_m - T + T \cdot \ln\left(\frac{T}{T_m}\right)\right] \quad (eq. 1)$$

que no necesita conocer la variación de entropía molar (ΔS), la cuál es más técnicamente difícil de determinar computacionalmente (aunque no imposible)⁶. De las simulaciones de DM se obtienen las entalpías absolutas de cada estado ($\langle H \rangle$) y con ellas se calculan las diferencias de entalpía (ΔH):

$$\Delta H_{NI}(T) = \langle H_I(T) \rangle - \langle H_N(T) \rangle \quad ; \quad \Delta H_{ID}(T) = \langle H_D(T) \rangle - \langle H_I(T) \rangle \quad (eqs. 2.1 y 2.2)$$

Y habiendo determinado las diferencias de entalpía (ΔH) a varias temperaturas, se puede extrapolar el valor a la temperatura media de desnaturalización (T_m) de cada transición ($\Delta H(T_m)$) y también determinar los calores específicos (ΔC_p), supuestos constantes a las temperaturas utilizadas:

$$\Delta H_{\boxed{NI \text{ ó } ID}}(T) = \Delta H(T_m) + \Delta C_p \cdot (T - T_m) \quad (eq. 3)$$

Ambos parámetros (tanto $\Delta H(T_m)$ como ΔC_p , para cada transición) son necesarios para la utilización de la ecuación de Gibbs-Helmholtz. Para la determinación de éstos es necesario conocer las T_m , que también aparecen en la ecuación y que para este trabajo se extraen de los datos de estabilidad de la apoflavodoxina ya publicados (método “semi-computacional”) (**Tabla 1**).

2.4. Formación de complejo ligando-proteína

Por otro lado, para la reacción de **unión no covalente de un ligando** a una proteína, también se establece un equilibrio entre los dos elementos libres y el complejo y se puede definir su energía libre de Gibbs:

$$N + FMN \leftrightarrow N - FMN \quad ; \quad K_b = \frac{\chi_{N-FMN}}{\chi_N \cdot \chi_{FMN}} \quad ; \quad \Delta G_b(T) = -R \cdot T \cdot \ln K_b(T) = \Delta H - T \cdot \Delta S$$

También es aplicable la ecuación de Gibbs-Helmholtz (eq. 1; $\Delta G_b(T)$) y la dependencia de la diferencia de entalpía con la temperatura (en lugar de T_m se utiliza una temperatura de referencia, en este caso 298 K (25 °C), a la que se han determinado los parámetros termodinámicos de unión (**Tabla 2**)) y su relación con el calor específico (eq. 3; $\Delta H_b(T)$ y $\Delta C_{p,b}$). En este caso, las entalpías absolutas obtenidas de las simulaciones de DM se relacionan por:

$$\langle H_{no-unido}(T) \rangle = \langle H_N(T) \rangle + \langle H_{FMN}(T) \rangle \quad (eq. 4)$$

$$\Delta H_b(T) = \langle H_{unido(N-FMN)}(T) \rangle - \langle H_{no-unido}(T) \rangle \quad (eq. 5)$$

Tabla 1. Datos termodinámicos de la apoflavodoxina de *Anabaena sp.*, determinados in vitro por experimentos de espectroscopía o por DSC (todos en MOPS 50 mM, pH 7). Se muestran también medias y errores estándar (Err. Est.) calculados del conjunto de datos de espectroscopía y de todos los datos in vitro en común (Exp.).

Origen	Ref. ó Dato	T_{m1} (K)	$\Delta H_{N \rightarrow I}$ (kcal/mol)	T_{m2} (K)	$\Delta H_{I \rightarrow D}$ (kcal/mol)	$\Delta C_{p,N \rightarrow I}$ (kcal/mol·K)	$\Delta C_{p,I \rightarrow D}$ (kcal/mol·K)	ΔC_p^a (kcal/mol·K)	$\Delta G_{298 K}^b$ (kcal/mol)
Espectroscopía	16	316,5	28,1	329,6	54,9				
	17	317,3	33,9	329	52,7				
	18	317,3	34,1	329	52,9				
	8	317,2	32,2	328,9	51				
	19	317,3	34,0	329,1	53,0				
	20	316	34,7	328	55,6				
	Media	316,9	32,8	328,9	53,4				
	Err. Est. ^c	0,2	1,0	0,2	0,7				
DSC (16)	Media	311,8	28,1	326,5	56	1,5 ^e	1,3 ^e	2,8 ^e	
	Err. Exp.	2,5	3,7	1,2	2,5	0,3 ^e	0,6 ^e	0,3 ^e	
Exp. ^d	Media	316,2	32,2	328,6	55,7	1,5	1,3	2,8	4,34 ^f
	Err. Est. ^c	0,8	1,1	0,4	0,7	0,3	0,6	0,3	1,8 ^f

$$^a\Delta C_p = \Delta C_{p,N \rightarrow I} + \Delta C_{p,I \rightarrow D}$$

$$^b\Delta G_{298 K} = \Delta G_{298 K, N \rightarrow I} + \Delta G_{298 K, I \rightarrow D}. \text{ Cada } \Delta G \text{ se calcula con la ecuación de Gibbs-Helmholtz (eq. 1).}$$

^cError estándar calculado como $ES = s/\sqrt{n}$, donde n es el número de datos y $s = \sqrt{\sum(x - \bar{x})^2 / (n - 1)}$ (\bar{x} media de los datos).

^dMedia y error estándar de todo el conjunto de datos experimentales, incluyendo los de espectroscopía y el de DSC.

^eEn Irún et al., 2001¹⁶ ofrecen un dato de $\Delta C_{p,N \rightarrow I}$ obtenido de la relación lineal de la diferencia de entalpía ($\Delta H_{N \rightarrow I}$) con la temperatura (eq. 3). En la misma figura (Figura 5 de la referencia), también están representados los datos de diferencia de entalpía global ($\Delta H_{N \rightarrow D}$) en función de la temperatura, y calculan de la misma manera (eq. 3) un valor de $\Delta C_{p,N \rightarrow D}$ diferente y menos preciso al que obtienen por ajustes de datos de DSC, ya que en la gráfica se observa un claro dato atípico. Tomando los datos de la recta de $\Delta H_{N \rightarrow D}$ vs. T y sin el dato atípico, se obtiene un valor más preciso (y más similar al obtenido por DSC) de $\Delta C_{p,N \rightarrow D}$ que es el que se recoge en esta tabla y que, junto con el valor de $\Delta C_{p,N \rightarrow I}$ ofrecido en la referencia, permite calcular $\Delta C_{p,I \rightarrow D}$ y su error estándar experimental (Err. Exp.).

^fPara el cálculo de la estabilidad experimental, se ha utilizado la ecuación integrada de Gibbs-Helmholtz (eq. 1) con las medias y errores estándar del conjunto de datos experimentales (Exp.), pero, en lugar de las medias y errores estándar de las T_m calculadas con todo el conjunto de datos in vitro (Exp.), se han utilizado las medias y errores estándar de las T_m calculadas sólo con el conjunto de datos de espectroscopía.

Tabla 2. Datos termodinámicos de la unión del FMN a la apoflavodoxina de *Anabaena* sp. en estado nativo determinados por experimentos in vitro (MOPS 50 mM, pH 7, 298 K).

	ΔG_b (kcal/mol)	K_b ($\times 10^9$ M)	ΔH_b^a (kcal/mol)	$\Delta C_{p,b}^a$ (kcal/mol·K)
Datos	-13,1 ²¹	3,8 ^{b,21}	(288 K) -5,8 ²¹	
	-13,3 ²²	5,7 ^{c,22}	-11,7 ²¹	-0,6 ²¹
	-12,3 ^{d,23}	1,1 ²³	(316 K) -16,6 ²¹	
Media	-12,9	3,5	-11,7	-0,6
Err. Est. ^e	0,3	1,3	0,6 ^{f,21}	0,02 ^{f,21}

^aMedidos en fosfato sódico 50 mM.

^bEl parámetro que aparece en la referencia es K_d , con el que se calcula K_b mediante la relación $K_b = 1/K_d$.

^c K_b se calcula con la ΔG_b reportada mediante la ecuación de Gibbs $K_b = \exp(-\Delta G_b/R \cdot T)$.

^d ΔG_b se calcula con la K_b reportada mediante la ecuación de Gibbs $\Delta G_b = -R \cdot T \cdot \ln(K_b)$.

^eError estándar calculado como en la **Tabla 1**, nota c.

^fErrores estándar que acompañan a los valores de los datos en las publicaciones originales.

3. Materiales y métodos

3.1. Estructuras del PDB

Se obtuvieron del Protein Data Bank (PDB) las estructuras correspondientes a los estados nativo e intermediario de las proteínas y la estructura de la holoproteína. La estructura del estado nativo (PDB ID: 1FTG⁷, **Figura 2 A**) está resuelta a 2 Angstroms de resolución y está completa, a excepción del residuo N-terminal, que no se ha añadido para la simulación. También se han eliminado las aguas cristalográficas y un anión sulfato. El intermediario tiene un conjunto de 20 estructuras obtenidas por RMN de un mutante puntual (PDB ID: 2KQU⁸, **Figura 2 D**), y para simularlas se ha eliminado el residuo N-terminal que no está en el nativo y se ha cambiado el aminoácido mutado de nuevo al original, para que las proteínas tengan el mismo número y tipo de átomos y por lo tanto sus sistemas sean comparables. La estructura de la holoproteína (PDB ID: 1FLV¹⁰, **Figura 2 A**) corresponde a la flavodoxina de *Anabaena* sp. cepa PCC 7120, pero tiene una similitud de secuencia del 100% con la de la cepa PCC 7119. También está resuelta a 2 Angstroms de resolución, sólo le falta el residuo N-terminal y se le han eliminado las aguas cristalográficas.

3.2. Estructuras del estado desnaturalizado

Para obtener un conjunto de más de 2000 estructuras que representen el estado desnaturalizado, se utilizó el servidor ProtSA⁹, que utiliza el algoritmo flexible-meccano²⁴ para generar conformaciones del esqueleto proteico (según los ángulos torsionales y los radios de giro de proteína desplegadas) y Scomp²⁵ para añadir las cadenas laterales (desde una librería de rotámeros y evitando impedimentos estéricos) a partir de una secuencia de aminoácidos, en este caso la de la estructura del PDB 1FTG. El

diámetro de las estructuras resultantes seguía una distribución estadística asimétrica debido a algunas estructuras con un diámetro excesivamente grande. Para obtener una distribución Gaussiana y también reducir el tamaño máximo de sistema necesario y con ello el gasto computacional, se estableció un valor límite de diámetro de 16 nm y se filtraron 1852 estructuras (**Figura 2 E**).

3.3. Tamaño mínimo de muestra del estado intermediario y desplegado

El número mínimo de trayectorias de los estados desplegado e intermediario que se deben simular y promediar para obtener unos valores de ΔH con un error similar o menor a los obtenidos de forma experimental, ya que los valores de entalpía absoluta siguen una distribución normal, se puede estimar con:

$$n = (z_{0,975} \cdot s_M / d)^2 \quad (eq. 6)$$

Siendo $z_{0,975} = 1,96$ el valor para una distribución estándar normal con un intervalo de confianza del 95% (bilateral), d el error máximo deseado (la mitad del error experimental de cada ΔH) y s_M la desviación estándar de una muestra de 40 trayectorias realizadas en un estudio preliminar. Se obtiene un tamaño de muestra de 100 trayectorias para los estados intermediario y desplegado.

3.4. Parámetros y etapas de simulación

De cada estructura o ensemble se realizaron simulaciones a tres **temperaturas de simulación**, 305, 320 y 335 K, que exploran un rango entre las que están las temperaturas medias de desnaturalización (T_m) de la apoflavodoxina determinadas de forma experimental: 316 y 329 K, aproximadamente. Con las entalpías absolutas ($\langle H(T) \rangle$) extraídas y promediadas se obtienen datos de $\Delta H(T)$ con los que se calculan los valores a las temperaturas de interés por extrapolación lineal y también el resto de parámetros utilizando las relaciones descritas por las ecuaciones presentadas. Para cada temperatura, para el estado nativo se realizan 40 **réplicas**, para cada uno de los 20 modelos del estado intermediario 5 réplicas (100 réplicas totales) y de las 1852 estructuras filtradas por tamaño del estado desplegado se seleccionan 100 al azar para simular independientemente. Las simulaciones relacionadas con el proceso de unión de FMN también se realizan a las mismas temperaturas y haciendo 20 réplicas de la holoproteína, del sistema con la apoflavodoxina y el FMN separados y del FMN sólo.

Las estructuras se procesaron (y las figuras se han preparado) con PyMOL 2.3.4, los estados de protonación de residuos de interés se determinaron con PROPKA¹⁴ a pH 7 a través del servidor PDB2PQR¹⁴ y todas las etapas de la simulación se realizaron con GROMACS 4.6.1²⁶⁻²⁸ en la plataforma de computación Memento, del Instituto de Biocomputación y Física de Sistemas Complejos (BIFI) de la Universidad de Zaragoza. La descripción del sistema se hace con el campo de fuerzas CHARMM22/CMAP²⁹ y se utiliza el modelo de agua TIP3P (*Transferable Intermolecular Potential with 3 Points*) de 3 puntos de interacción³⁰. Además, para la obtención de los parámetros del FMN, se utiliza Amber³¹ y el campo de fuerzas general de Amber (*General Amber Force Field*, GAFF)³², asignando las cargas con el método del RESP (*Restrained Electrostatic Potential*)³³ después de computarlas *ab initio* con Gaussian09³⁴ al nivel teórico 6-31G*³⁵ y utilizando el método de Hartree-Fock (HF)³⁶.

Las estructuras cristalinas procesadas se transformaron a un formato adecuado para GROMACS según el campo de fuerzas elegido para su **preparación** para la simulación. Primero, se minimizaron en vacío con el algoritmo de descenso de máxima pendiente (*steepest descent*) y valores límite simples para interacciones a larga distancia. Luego, la estructura nativa se introdujo en una caja dodecahédrica rómbica de 18 nm de lado, mientras que para el resto de estructuras se utilizaron en un primer momento sistemas más grandes para, tras solvatar, igualar el número de moléculas de agua en todos los sistemas y que tuvieran el mismo número de átomos. Para neutralizar las cargas negativas de las estructuras y también del FMN, se introducen los cationes de sodio necesarios para ello, sin establecer una concentración iónica, y finalmente se vuelve a realizar una minimización más larga, utilizando en este caso los algoritmos PME y LINCS. Además, para reajustar el volumen de los sistemas en los que se ha reducido el número de moléculas de agua, se realiza un primer paso de DM propiamente dicho, con el integrador Leapfrog y un paso de tiempo de 1 fs durante 200 ps. También se utilizan PME y LINCS, y se introduce el baróstato Berendsen y el termostato Bussi-Donadio-Parrinello a 0 K.

El **calentamiento** se realiza en 7 pasos de 50 ps (350 ps en total) en los que se sube la temperatura de forma escalonada hasta la deseada y utilizando las restricciones de posición de los átomos de los archivos itp, aunque sin utilizar el baróstato.

En el primer paso de **equilibrado** de 150 ps se continúa la simulación en las mismas condiciones que el último paso de calentamiento para estabilizar el sistema a la temperatura deseada, pero eliminando las restricciones de posición. En el segundo y el tercero, de 250 ps cada uno con un paso de tiempo mayor de 2 fs, se reintroduce el baróstato Berendsen y se cambia por el Parrinello-Rahman isotrópico, respectivamente; y también se separa el funcionamiento del termostato entre la proteína (y el FMN) y el resto del sistema, para controlar la temperatura de forma separada en ellos.

La **producción** es de 2 ns también con un paso de tiempo de 2 fs, recogiendo y extrayendo la información energética cada 100 fs, para obtener promedios a lo largo del tiempo con un bajo error. Se realiza en las mismas condiciones que el último paso del equilibrado: utilizando los algoritmos PME y LINCS, el termostato Bussi-Donadio-Parrinello y el baróstato Parrinello-Rahman isotrópico. Para evitar desviaciones de los valores de energía asociados a la utilización de computación paralela, se utiliza el balanceo dinámico de la carga computacional, que redistribuye continuamente los cálculos a realizar entre los núcleos utilizados para evitar que los más lentos hagan que el resto estén mucho tiempo esperando a que terminen; y no se separan los cálculos del algoritmo PME del resto de cálculos de interacciones en distintos núcleos. El análisis de datos se ha hecho con los propios módulos de GROMACS, Microsoft Excel y Origin 2018.

4. Resultados y discusión

4.1. Estabilidad de la apoflavodoxina

Con las entalpías absolutas de los sistemas de cada estado ($\langle H(T) \rangle$) obtenidas de las simulaciones, se calculan las diferencias de entalpía entre los estados (eqs 2.1 y 2.2) a cada una de las tres temperaturas ($\Delta H(T)$), y con la dependencia lineal de éstas con la temperatura (eq. 3) se obtienen las diferencias de entalpía a las temperaturas medias de desnaturalización (T_m) y las diferencias de calor específico (ΔC_p)

para cada transición. Finalmente, con las diferencias de entalpía y las de calor específico entre cada estado calculadas y las temperaturas medias de desnaturalización experimentales, se calcula la estabilidad como la diferencia de energía libre ($\Delta G(T)$) según la ecuación de Gibbs-Helmholtz (eq. 1).

De cada trayectoria de DM, se obtiene el valor de entalpía absoluta como un promedio a lo largo del tiempo y luego se promedian todas las réplicas realizadas de cada estado, obteniendo así la media del valor de entalpía absoluta de cada estado (y su error estándar), que es el que se utiliza para los cálculos. Para los tres estados, los valores de $\langle H \rangle$ son del orden de 10^6 kcal/mol, mientras que los valores experimentales de ΔH que se quieren calcular por diferencia son del orden de 10^1 kcal/mol. Por esta razón, se requiere una gran precisión del campo de fuerzas, que cuantifique correctamente las diferencias debidas al establecimiento de interacciones distintas en los tres estados de plegamiento. Debido al gran número de datos con los que se realiza el promediado temporal (20.001 puntos de energías extraídas cada 100 fs), los errores estándar de los múltiples valores de $\langle H \rangle$ van de tan sólo 1,8 a 29 kcal/mol, mientras que el error estándar de los promedios de $\langle H \rangle$ de los distintos estados va de 1,4 a un máximo de 5,7 kcal/mol, por lo que, teniendo en cuenta que los promedios también son del orden de 10^6 kcal/mol, se obtiene una gran precisión en ambos casos.

El valor calculado de $\Delta H_{N \rightarrow I}(T_{m1})$ es $36,5 \pm 6,0$ kcal/mol y el de $\Delta H_{I \rightarrow D}(T_{m2})$ es $48,4 \pm 4,1$ kcal/mol (media \pm err. est.), que coinciden razonablemente con los datos experimentales promediados de $32,2 \pm 1,1$ y $55,7 \pm 0,7$ kcal/mol, respectivamente, dentro del margen de error (Tabla 3). También a partir de los datos de ΔH a varias temperaturas se pueden calcular los ΔC_p mediante la eq. 3, obteniendo unos valores de $\Delta C_{p,N \rightarrow I}$ de $1,5 \pm 0,09$ kcal/mol y de $\Delta C_{p,I \rightarrow D}$ de $1 \pm 0,02$ kcal/mol y que también concuerdan con los experimentales de $1,5 \pm 0,3$ y \pm kcal/mol¹⁶, respectivamente, y más aún si se compara la ΔC_p global, siendo $2,5 \pm 0,1$ kcal/mol la calculada y $2,8 \pm 0,3$ kcal/mol la experimental (Tabla 3).

Tabla 3. Media y error estándar de los datos termodinámicos de la apoflavodoxina de *Anabaena* sp. determinados *in vitro* (Exp.) (Tabla 1) y calculados por Dinámica Molecular (DM).

Origen	Ref. ó Dato	T_{m1} (K)	$\Delta H_{N \rightarrow I}$ (kcal/mol)	T_{m2} (K)	$\Delta H_{I \rightarrow D}$ (kcal/mol)	$\Delta C_{p,N \rightarrow I}$ (kcal/mol·K)	$\Delta C_{p,I \rightarrow D}$ (kcal/mol·K)	ΔC_p (kcal/mol·K)	$\Delta G_{298 K}$ (kcal/mol)
Exp.	Media	316,9 ^a	32,2	328,9 ^a	55,7	1,5	1,3	2,8	4,34
	Err. Est.	0,2 ^a	1,1	0,2 ^a	0,7	0,3	0,6	0,3	1,8
DM	Calc.	-	36,5	-	48,4	1,5	1	2,5	4,36
	Err. Est.	-	6,0	-	4,1	0,09	0,02	0,1	1,4

^aLa media y error estándar de las T_m utilizadas para los cálculos (también para los cálculos con los datos obtenidos por DM) son las del conjunto de datos determinados por técnicas de espectroscopía (Tabla 1).

Con los datos calculados, se puede aplicar la ecuación integrada de Gibbs-Helmholtz (eq. 1) para obtener la estabilidad (ΔG) de la proteína. A la temperatura ambiente de 298 K (25 °C) a la que se refieren las mediciones de estabilidad *in vitro*, la estabilidad calculada de la proteína es de $4,36 \pm 1,4$ kcal/mol y la calculada a partir de los datos experimentales es $4,34 \pm 1,8$ kcal/mol, que coinciden de manera sobresaliente dentro del margen de error (Tabla 3). Para el cálculo de ambos valores se utilizan las

mismas medias y errores estándar de las T_m del conjunto de datos de técnicas de espectroscopía (Tabla 3, nota a), que al ser varios datos (y además ser más similares entre ellos que con el dato de DSC) permiten obtener un error estándar pequeño con el que se calculan las $\Delta G_{298 K}$ (eq. 1) con un error asumible. Representando las curvas de estabilidad de la proteína resultantes (Figura 4 A), la calculada con los datos de DM se superpone casi por completo a la calculada con los datos de experimentos *in vitro* (Tabla 3) en el rango de temperaturas representado (de 220 a 400 K). Además, con todos los datos, también se puede reconstruir el perfil de exceso de calor específico (termograma de DSC), que se corresponde adecuadamente al experimental (Figura 4 B). Por otro lado, también se pueden obtener las fracciones proteicas de los distintos estados en el rango de temperaturas en el que se produce el desplegamiento (de 300 a 350 K) utilizando tanto los datos obtenidos de DM (Figura 4 C) como la media de todos los datos experimentales (Figura 4 D) de la Tabla 3, y se observa como ambas representaciones coinciden de manera excelente con la experimental de Irún *et al.*, 2001¹⁶ y, por extensión, entre ellas.

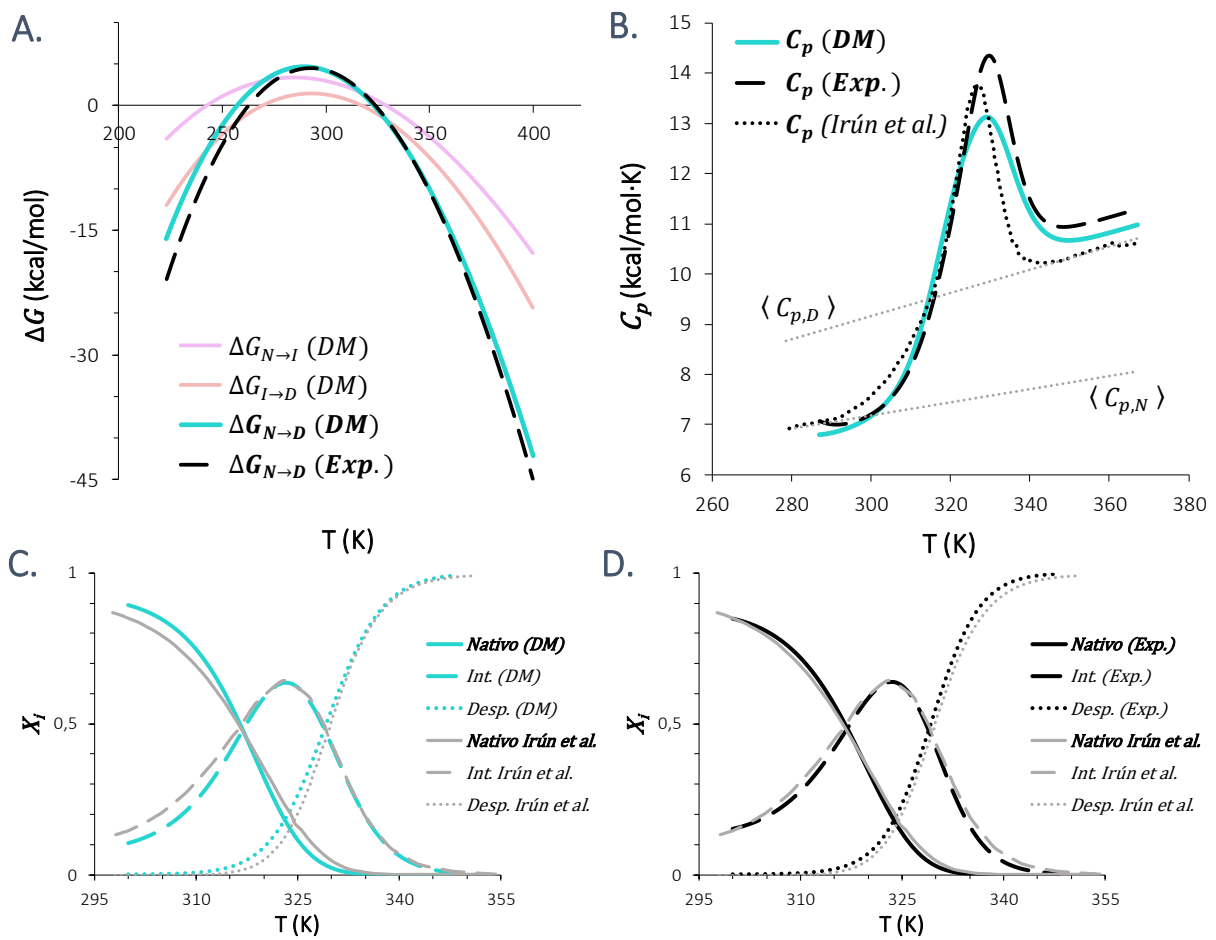


Figura 4. **A.** Curvas de estabilidad de la apoflavodoxina de *Anabaena sp.* Se representan las curvas de energía libre de las dos transiciones de la ruta de desplegamiento térmico, calculadas según la ecuación de Gibbs-Helmholtz (eq. 1) con los datos obtenidos de DM ($\Delta G_{N \rightarrow I}$ (DM) y $\Delta G_{I \rightarrow D}$ (DM)), y la suma de ambas ($\Delta G_{N \rightarrow D}$ (DM)). También se han calculado las curvas de energía libre de las dos transiciones a partir de las medias de todos los datos obtenidos de experimentos *in vitro* (no mostradas; utilizando los datos de la Tabla 3) y se representa la suma de ambas ($\Delta G_{N \rightarrow D}$ (Exp.)). **B.** Perfiles de exceso de calor específico de la apoflavodoxina de *Anabaena sp.* Se representan también con líneas de puntos grises las rectas correspondientes a los C_p del estado nativo ($\langle C_{p,N} \rangle$) y del desnaturalizado ($\langle C_{p,D} \rangle$) que aparecen en el perfil de Irún *et al.*, 2001¹⁶. Los perfiles se calculan con los datos de DM y experimentales (Exp.) de la Tabla 3, utilizando la ecuación descrita en Sánchez-Ruiz, 1995³⁷ sumándole el término de línea base $B_0 + B_1 \cdot T$, siendo B_0 y B_1 las pendientes de las rectas de $\langle C_{p,N} \rangle$ y $\langle C_{p,D} \rangle$ de Irún *et al.*, 2001¹⁶. **C. y D.** Perfiles de la fracción proteica (X_i) de cada uno de los estados, de 300 a 350 K, calculados según lo descrito en Sánchez-Ruiz, 1995³⁷ y comparando el perfil obtenido con los datos de DM (C.) o de la media de datos experimentales (Exp.) (D.) de la Tabla 3 con el perfil obtenido por ajuste de datos de técnicas de espectroscopía de Irún *et al.*, 2001¹⁶, respectivamente.

Cada simulación ejecutada dura un total de 3 ns (teniendo en cuenta todos los pasos), por lo que habiendo realizado un total 240 de para obtener todos los datos, el tiempo total de simulación es 0,72 μ s. Este tiempo de simulación es fácilmente alcanzable con la capacidad computacional actual (teniendo en cuenta los tamaños de los sistemas utilizados) y confirma que este método de cálculo de la estabilidad proteica se puede aplicar a cualquier proteína con un gasto computacional asequible. Las conformaciones intermediarias y, sobre todo, de las desplegadas, acaparan la mayor parte de las simulaciones, ya que sus *ensembles* tienen una mayor diversidad conformacional y por lo tanto necesitan una mayor exploración.

Por otro lado, la precisión de los datos alcanzada indica que el campo de fuerzas y el modelo de agua utilizados son adecuados para este método, como ya se observó en Galano-Frutos *et al.*, 2019⁶, y también indica que los conjuntos de estructuras utilizadas para simular los estados intermediario y desnaturalizado los representan adecuadamente. Las energías que se extraen de las simulaciones de estos estados dan lugar a datos termodinámicos razonablemente coincidentes (dentro del margen de error) con los determinados por experimentos *in vitro*, por lo que se puede afirmar que las estructuras simuladas y las posiciones de los átomos y las interacciones presentes en ellas son fieles a la realidad y reproducen correctamente su energética. Es de especial interés la confirmación de que ProtSA⁹ genera *ensembles* de estados desplegados representativos, ya que se basa en datos *in vitro* de proteínas con estados desnaturalizados completamente desplegados para hacerlo (en principio, no serviría para proteínas que conservan algo de estructura en este estado).

Sin embargo, un problema común de los campos de fuerza y modelos de agua comunes es que recrean adecuadamente el comportamiento de los estados plegados, para los cuáles están más optimizados, mientras que la simulación de estados desplegados es más problemática³⁸. Las partes de las proteínas o las proteínas desplegadas no son solvatadas correctamente y como consecuencia tienden a la compactación³⁸, potencialmente dejando de ser representativas de sus estados. Aun así, los cortos tiempos de simulación utilizados probablemente no permiten que ésto ocurra. Para comprobarlo, se han extraído los radios de giro (R_g) de todas las estructuras simuladas de los tres estados y a las tres temperaturas, tanto de las estructuras de partida (PDB o ProtSA) como a distintos tiempos de la etapa de producción (**Figura 5**), y se ha analizado su variación a lo largo de la simulación. Como es de esperar, las variaciones de los radios de giro de las estructuras del estado nativo no varían significativamente (p -valor $>0,05$) y además se determina que las del estado desplegado tampoco (**Figura 5**). Pese a esto, en el análisis estadístico se obtiene que los R_g de las estructuras del estado intermediario sí varían significativamente (p -valor $<0,05$). Aunque, si se comparan la primera y la última media de los R_g de los intermediarios a cada temperatura con las correspondientes del estado desplegado (**Figura 5**), se puede ver como la variación en valor absoluto es mucho más pequeña en el estado intermediario (variaciones de 0,05, 0,07 y 0,06 a cada temperatura, respectivamente) que en el desplegado (de 0,1, 0,19 y 0,19, respectivamente), mientras que en el desplegado la mayor variación no resulta estadísticamente significativa debido al amplio rango de valores de R_g . Por otro lado, en Galano-Frutos *et al.*, 2019⁶ también se determinó que no hay relación entre el R_g y la diferencia de entalpía calculada resultante, por lo que los valores de parámetros termodinámicos obtenidos no se ven afectados por la existencia de una compactación significativa y este resultado no afecta a la validez del método propuesto.

Finalmente, atendiendo a la hipótesis de ergodicidad, las simulaciones cortas de varias conformaciones representativas de los *ensembles* de los estados intermediario o desplegado se podrían sustituir por unas pocas simulaciones largas (por ejemplo, con producciones de 100 ns) de unas pocas

conformaciones de cada uno, potencialmente obteniendo unos promedios comparables. En tal caso, además de que el tiempo de simulación total aumentaría (101 ns por simulación y, por ejemplo, 3 simulaciones por estado para promediar: 0,91 μ s totales), en las trayectorias tan largas probablemente sí que daría tiempo a que se produjera una compactación significativa y, potencialmente, variaciones en la energía de los sistemas, que en última instancia afectarían a los resultados obtenidos y su precisión, como se demostró en Galano-Frutos *et al.*, 2019⁶.

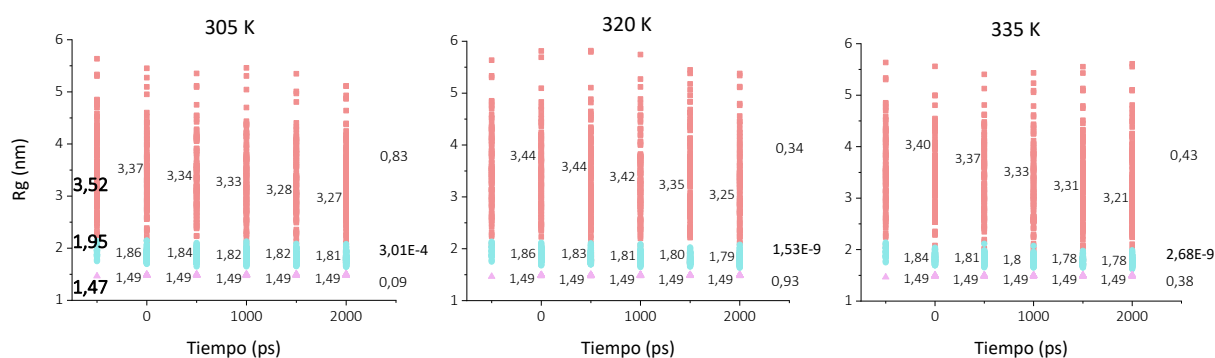


Figura 5. Evolución de los Rg de las estructuras en simulación de los distintos estados, desde la estructura inicial (primera línea de datos a -500 ps) y en varios puntos hasta el final de la producción, a las 3 temperaturas. Los cuadrados rosas corresponden al estado desplegado, los círculos azules al intermediario y los triángulos lavanda al nativo. Al lado de cada conjunto de datos de cada estado a cada tiempo se muestra la media (los datos de la estructura inicial a -500 ps son iguales para las tres gráficas y se muestran una vez en la primera resaltados en negrita). En la parte de la derecha de cada una de las gráficas, se muestra el p-valor obtenido del análisis de varianza (One-Way ANOVA) de comparación de las medias de los valores de Rg de cada estado a los distintos tiempos (excluyendo la estructura inicial), con un nivel de confianza de 95%, resaltando en negrita los p-valor menores a 0,05.

4.2. Unión del FMN

Se ha utilizado el mismo procedimiento para obtener datos de diferencia de entalpía ($\Delta H_b(T)$), diferencia de calor específico ($\Delta C_{p,b}$) y diferencia de energía libre de Gibbs ($\Delta G_b(T)$) de la unión del FMN a la apoflavodoxina de *Anabaena* sp. a través de simulaciones de Dinámica Molecular (**Tabla 4**). En este caso, para la obtención de $\Delta H_b(T)$ se debe restar la entalpía absoluta del estado unido ($\langle H_{unido(N-FMN)}(T) \rangle$) menos la del estado no-unido ($\langle H_{no-unido}(T) \rangle$) (eq. 5). Para simular el estado no-unido, en el que el FMN y la apoflavodoxina nativa están separados, se han utilizado dos representaciones distintas (**Figura 1 B**), y posteriormente se han realizado los cálculos también independientemente para cada una de ellas para su comparación. En la representación 1 (Rep. 1 en **Tabla 4**), se utilizan los datos extraídos de las simulaciones de la apoproteína en su estado nativo de la primera parte del trabajo y se complementan con 20 simulaciones del FMN sólo (en un sistema pequeño con las moléculas de agua necesarias para solvatar y utilizar PBC sin aparición de artefactos), para utilizar la suma de los datos de ambos conjuntos de simulaciones como los datos de energía del estado no-unido (eq. 4) (**Figura 1 B**). En la representación 2 (Rep. 2 en **Tabla 4**), se sitúan la apoproteína y el FMN en un mismo sistema (separados lo suficiente para evitar que interaccionen) (**Figura 1 B**), en el que se introducen el número de aguas resultado de la suma de las que hay presentes en los sistemas del estado plegado y del FMN, respectivamente, de la representación 1 para que el número y tipo de átomos en ambas representaciones sea el mismo y sus energéticas comparables. El estado unido se simula utilizando la estructura de la holoproteína (**Figura 1 B**) y con el mismo número de aguas que el sistema de la representación 2, por la misma razón.

Tabla 4. Media y error estándar de los datos termodinámicos de la unión del FMN a la apoflavodoxina de *Anabaena sp.* determinados *in vitro* (Exp.) (de **Tabla 2**, todos a 298 K) y calculados por Dinámica Molecular (DM) con las dos representaciones (Rep.) utilizadas del estado no-unido (apoproteína y FMN en sistemas diferentes (1) o en el mismo pero separados (2)).

		ΔG_b (kcal/mol)	K_b ($\times 10^9$ M)	ΔH_b (kcal/mol)	$\Delta C_{p,b}$ (kcal/mol·K)
Exp.	Media	-12,9	3,5	-11,7	-0,6
	Err. Est.	0,3	1,3	0,6	0,02
DM	Media			-4,18	0,91
	Err. Est.			9	0,20
	Media			10,74	0,38
	Err. Est.			8	0,03

El valor calculado resultante de $\Delta H_{b,298 K}$ es $-4,18 \pm 9$ kcal/mol utilizando la representación 1 y $10,74 \pm 8$ kcal/mol utilizando la 2 (**Tabla 4**), ambos con un error estándar asociado grande, del mismo orden que el propio valor. El valor determinado por experimentos *in vitro* es $-11,7 \pm 0,6$ kcal/mol, por lo que sólo con la representación 1 se obtiene un valor que es concordante, dentro del gran margen de error, con el experimental. Sin embargo, comparando la relación lineal con la temperatura de las diferencias de entalpía experimentales y calculadas (**Figura 6 A**) se ve claramente que, en realidad, ninguna de las dos representaciones reproduce correctamente la energética del fenómeno, ya que las dependencias calculadas tienen una pendiente de signo contrario a la dependencia real de la diferencia de entalpía de unión con la temperatura. Las pendientes de estas rectas corresponden al calor específico ($\Delta C_{p,b}$), que en la reacción de unión de un ligando a una proteína es casi siempre negativo³⁹, como refleja el valor determinado *in vitro* de $-0,6 \pm 0,02$ kcal/mol·K. El valor correspondiente a la representación 1 resulta ser $0,91 \pm 0,20$ kcal/mol·K y con la 2 es $0,38 \pm 0,03$ kcal/mol·K (**Tabla 4**), ambos valores de signo positivo y muy alejados del experimental.

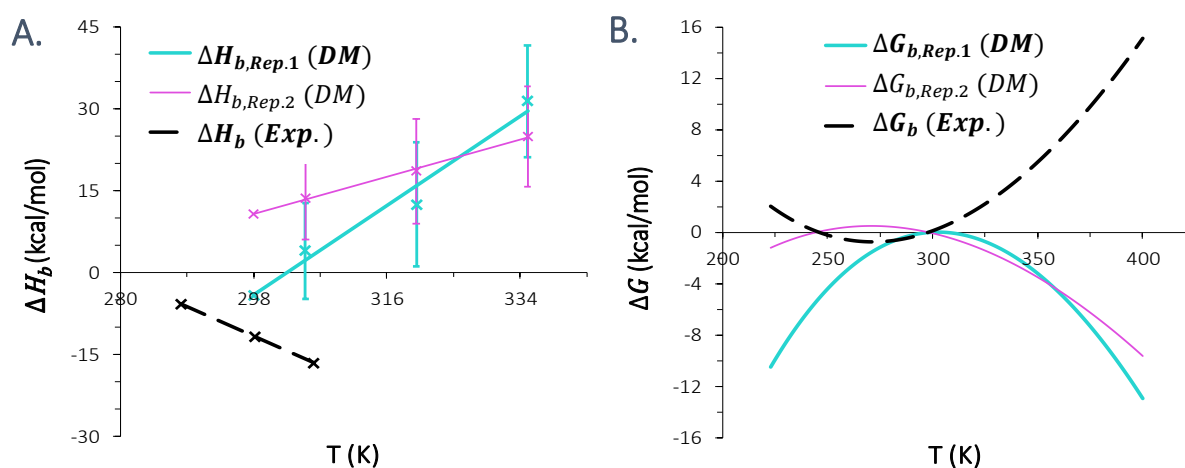


Figura 6. **A.** Relación lineal con la temperatura de las diferencias de entalpía de unión (ΔH_b) del FMN a la apoflavodoxina de *Anabaena sp.* experimentales (Exp.) (**Tabla 4**) y obtenidas con Dinámica Molecular (DM), con las dos representaciones (Rep.) utilizadas del estado no-unido. El cálculo de las diferencias de entalpía con DM para cada representación se realiza a tres temperaturas (305, 320 y 335 K) y se representan sus barras de error y también el valor correspondiente extrapolado a 298 K. **B.** Representación de la estabilidad de unión (ΔG_b) del FMN a la apoflavodoxina de *Anabaena sp.* en función de la temperatura. Las curvas están calculadas con la ecuación de Gibbs-Helmholtz (eq. 1) utilizando los datos experimentales (Exp.) o los obtenidos de DM para las dos representaciones (Rep.) (**Tabla 4**, datos de ΔH_b y $\Delta C_{p,b}$).

Al utilizar la ecuación integrada de Gibbs-Helmholtz (eq. 1) para calcular la estabilidad de la unión en función de la temperatura ($\Delta G_b(T)$) y representar las curvas resultantes (**Figura 6 B**), se termina de confirmar que las energéticas simuladas con este método con cualquiera de las dos representaciones no son fiables, ya que las curvas son completamente diferentes a la calculada con datos experimentales. Para intentar obtener datos que potencialmente se ajusten más a la realidad, se podría repetir el procedimiento disminuyendo el exceso de aguas de los sistemas (que al no necesitarse la simulación de estados desplegados no son necesarias en tanta cantidad y pueden enmascarar las pequeñas diferencias de interacciones entre los estados no-unido y unido), utilizando parametrizaciones distintas del FMN y/o también con otros campos de fuerza y modelos de agua.

5. Conclusiones

Los resultados de este trabajo, junto con los obtenidos en Galano-Frutos *et al.*, 2019⁶, confirman que la utilización de simulaciones de Dinámica Molecular para el cálculo de la estabilidad de proteínas de 2 o 3 estados de desplegamiento es un método riguroso, preciso, asequible y accesible. Se basa en la definición de la diferencia de energía libre (ΔG) de la ecuación de Gibbs-Helmholtz (eq. 1), que utiliza las diferencias de entalpía (ΔH) calculadas con los datos extraídos de las simulaciones, las diferencias de calor específico (ΔC_p , calculadas a partir de datos de diferencias de entalpía a varias temperaturas) y las temperaturas medias de desnaturalización (T_m), las cuales son determinables de manera sencilla *in vitro* (método “semi-computacional”), mientras que no necesita conocer la diferencia de entropía (ΔS), que es más técnicamente difícil de calcular computacionalmente. Se han reproducido los parámetros termodinámicos del desplegamiento térmico de la apoflavodoxina de *Anabaena* sp. PCC 7119 con una excelente coincidencia, dentro del margen de error, con los parámetros determinados por experimentos *in vitro*. Además, la similitud de los datos obtenidos con los experimentales que se querían recrear también contribuye a demostrar la precisión del campo de fuerza y modelo molecular del agua utilizado (CHARMM22/CAMP²⁹ y TIP3P³⁰), la correcta representación de las conformaciones del estado intermediario de la apoflavodoxina de *Anabaena* sp. por parte de un mutante de un aminoácido diseñado⁸ y también la correcta representación del estado desplegado de las proteínas que suponen los *ensembles* generados con ProtSA⁹, que utiliza datos experimentales de estados desnaturalizados completamente desplegados de proteínas para generarlos. Sin embargo, al extender la aplicación del método al cálculo de la estabilidad de unión de un ligando (el FMN a la misma proteína), no se ha logrado obtener unos buenos resultados. Probablemente, esto se debe a que la precisión necesaria para ello es un orden de magnitud mayor que la obtenida, debido a la magnitud relativa de las entalpías de plegamiento y de unión de la proteína estudiada. En un futuro, es posible que la precisión del método pueda aumentar gracias al desarrollo de campos de fuerza y modelos moleculares de agua cada vez más rigurosos³⁸ y también se va a completar su validación extendiéndolo a más proteínas, además de intentar nuevamente recrear la energética de unión de ligando correctamente aplicando las soluciones propuestas. En su estado actual, el método constituye una forma precisa y asequible de calcular la estabilidad proteica *in silico*, con importantes implicaciones en el campo de las proteínas terapéuticas y comerciales.

Conclusions

This work's results, along with those obtained in Galano-Frutos *et al.*, 2019⁶, confirm that using Molecular Dynamics (MD) simulations for the calculation of the stability of 2-state or 3-state proteins is a rigorous, precise, achievable and accessible method. It exploits the definition of the free energy difference (ΔG) given by the Gibbs-Helmholtz equation (eq. 1), which utilizes the enthalpy differences (ΔH) calculated with the data extracted from the simulations, the specific heat differences (ΔC_p , calculated using the enthalpy differences data at various temperatures) and the melting temperatures (T_m), easily measured *in vitro* ("semi-computational" method), but without the need to know the entropy, harder to calculate computationally. The thermodynamical parameters of the thermal unfolding of the apoflavodoxin from *Anabaena* sp. PCC 7119 have been reproduced with significant precision and they exhibit an excellent agreement with those determined *in vitro*, within the margin of error. Moreover, the similarities of the gathered data with those which we wished to recreate, also assist to demonstrate the precision of the force field and water model used (CHARMM22/CMAP²⁹ and TIP3P³⁰), the adequate representation of the conformations of the apoflavodoxin's intermediate state by a single-point, designed protein mutant⁸ and, finally, the correct representation of the denatured state of proteins by the *ensembles* generated by ProtSA⁹, which is built upon experimental data of completely unfolded, denatured states of proteins. However, extending the procedure to the calculation of the binding of a ligand to a protein (FMN to apoflavodoxin) has not yielded the expected results. Possibly, this is due to the need of a 10-fold increase in the accuracy of the calculations (compared to what was achieved), considering the relative magnitude of the folding and binding enthalpies of the protein studied here. In the future, the precision of this method could improve following the development of ever more rigorous force fields and water models³⁸, while it is also going to be applied to more proteins to complete its validation, as well as to the ligand binding calculation, again, with the introduction of the proposed adjustments. In its present state, this method constitutes a precise and easily attainable procedure to calculate the stability of proteins *in silico*, with significant implications for therapeutic and commercial proteins.

Bibliografía

- (1) *Molecular Modeling of Proteins*, 2nd ed.; Kukol, A., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2015; Vol. 1215. <https://doi.org/10.1007/978-1-4939-1465-4>.
- (2) Sancho, J. The Stability of 2-State, 3-State and More-State Proteins from Simple Spectroscopic Techniques... plus the Structure of the Equilibrium Intermediates at the Same Time. *Arch. Biochem. Biophys.* **2013**, *531* (1–2), 4–13. <https://doi.org/10.1016/j.abb.2012.10.014>.
- (3) Bedouelle, H. Principles and Equations for Measuring and Interpreting Protein Stability: From Monomer to Tetramer. *Biochimie* **2016**, *121*, 29–37. <https://doi.org/10.1016/j.biochi.2015.11.013>.
- (4) *Computational Tools for Chemical Biology*, 1st ed.; Martín-Santamaría, S., Ed.; Chemical Biology; Royal Society of Chemistry: Cambridge, UK, 2017. <https://doi.org/10.1039/9781788010139>.
- (5) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* (80-.). **2012**, *338* (6110), 1042–1046. <https://doi.org/10.1126/science.1219021>.
- (6) Galano-Frutos, J. J.; Sancho, J. Accurate Calculation of Barnase and SNase Folding Energetics Using Short Molecular Dynamics Simulations and an Atomistic Model of the Unfolded Ensemble: Evaluation of Force Fields and Water Models. *J. Chem. Inf. Model.* **2019**, *59* (10), 4350–4360. <https://doi.org/10.1021/acs.jcim.9b00430>.
- (7) Genzor, C. G.; Perales-Alcón, A.; Sancho, J.; Romero, A. Closure of a Tyrosine/Tryptophan Aromatic Gate Leads to a Compact Fold in Apo Flavodoxin. *Nat. Struct. Mol. Biol.* **1996**, *3* (4), 329–332. <https://doi.org/10.1038/nsb0496-329>.
- (8) Ayuso-Tejedor, S.; Angarica, V. E.; Bueno, M.; Campos, L. A.; Abián, O.; Bernadó, P.; Sancho, J.; Jiménez, M. A. Design and Structure of an Equilibrium Protein Folding Intermediate: A Hint into Dynamical Regions of Proteins. *J. Mol. Biol.* **2010**, *400* (4), 922–934. <https://doi.org/10.1016/j.jmb.2010.05.050>.
- (9) Estrada, J.; Bernadó, P.; Blackledge, M.; Sancho, J. ProtSA: A Web Application for Calculating Sequence Specific Protein Solvent Accessibilities in the Unfolded Ensemble. *BMC Bioinformatics* **2009**, *10* (1), 104. <https://doi.org/10.1186/1471-2105-10-104>.
- (10) Rao, S. T.; Shaffie, F.; Yu, C.; Satyshur, K. A.; Stockman, B. J.; Markley, J. L.; Sundaralingam, M. Structure of the Oxidized Long-Chain Flavodoxin from Anabaena 7120 at 2 Å Resolution. *Protein Sci.* **1992**, *1* (11), 1413–1427. <https://doi.org/10.1002/pro.5560011103>.
- (11) Sancho, J. Flavodoxins: Sequence, Folding, Binding, Function and Beyond. *Cell. Mol. Life Sci.* **2006**, *63* (7–8), 855–864. <https://doi.org/10.1007/s00018-005-5514-4>.
- (12) García-Fandiño, R.; Bernadó, P.; Ayuso-Tejedor, S.; Sancho, J.; Orozco, M. Defining the Nature of Thermal Intermediate in 3 State Folding Proteins: Apoflavodoxin, a Study Case. *PLoS Comput. Biol.* **2012**, *8* (8), e1002647. <https://doi.org/10.1371/journal.pcbi.1002647>.
- (13) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1* (1), 1–28. <https://doi.org/10.33011/livecoms.1.1.5957>.
- (14) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), W665–W667. <https://doi.org/10.1093/nar/gkh381>.
- (15) Sherwood, D.; Dalby, P. *Modern Thermodynamics for Chemists and Biochemists*, 1st ed.; Oxford

University Press: Oxford, UK, 2018; Vol. 1.
<https://doi.org/10.1093/oso/9780198782957.001.0001>.

- (16) Irún, M. P.; Garcia-Mira, M. M.; Sanchez-Ruiz, J. M.; Sancho, J. Native Hydrogen Bonds in a Molten Globule: The Apoflavodoxin Thermal Intermediate. *J. Mol. Biol.* **2001**, *306* (4), 877–888. <https://doi.org/10.1006/jmbi.2001.4436>.
- (17) Campos, L. A.; Bueno, M.; Lopez-Llano, J.; Jiménez, M. Á.; Sancho, J. Structure of Stable Protein Folding Intermediates by Equilibrium ϕ -Analysis: The Apoflavodoxin Thermal Intermediate. *J. Mol. Biol.* **2004**, *344* (1), 239–255. <https://doi.org/10.1016/j.jmb.2004.08.081>.
- (18) Bueno, M.; Cremades, N.; Neira, J. L.; Sancho, J. Filling Small, Empty Protein Cavities: Structural and Energetic Consequences. *J. Mol. Biol.* **2006**, *358* (3), 701–712. <https://doi.org/10.1016/j.jmb.2006.02.060>.
- (19) Ayuso-Tejedor, S.; García-Fandiño, R.; Orozco, M.; Sancho, J.; Bernadó, P. Structural Analysis of an Equilibrium Folding Intermediate in the Apoflavodoxin Native Ensemble by Small-Angle X-Ray Scattering. *J. Mol. Biol.* **2011**, *406* (4), 604–619. <https://doi.org/10.1016/j.jmb.2010.12.027>.
- (20) Lamazares, E.; Clemente, I.; Bueno, M.; Velázquez-Campoy, A.; Sancho, J. Rational Stabilization of Complex Proteins: A Divide and Combine Approach. *Sci. Rep.* **2015**, *5*. <https://doi.org/10.1038/srep09129>.
- (21) Lostao, A.; El Harrou, M.; Daoudi, F.; Romero, A.; Parody-Morreale, A.; Sancho, J. Dissecting the Energetics of the Apoflavodoxin-FMN Complex. *J. Biol. Chem.* **2000**, *275* (13), 9518–9526. <https://doi.org/10.1074/jbc.275.13.9518>.
- (22) Campos, L. A.; Sancho, J. Native-Specific Stabilization of Flavodoxin by the FMN Cofactor: Structural and Thermodynamical Explanation. *Proteins Struct. Funct. Bioinforma.* **2006**, *63* (3), 581–594. <https://doi.org/10.1002/prot.20855>.
- (23) Lamazares, E.; Vega, S.; Ferreira, P.; Medina, M.; Galano-Frutos, J. J.; Martínez-Júlvez, M.; Velázquez-Campoy, A.; Sancho, J. Direct Examination of the Relevance for Folding, Binding and Electron Transfer of a Conserved Protein Folding Intermediate. *Phys. Chem. Chem. Phys.* **2017**, *19* (29), 19021–19031. <https://doi.org/10.1039/C7CP02606D>.
- (24) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J. R.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. Flexible-Meccano: A Tool for the Generation of Explicit Ensemble Descriptions of Intrinsically Disordered Proteins and Their Associated Experimental Observables. *Bioinformatics* **2012**, *28* (11), 1463–1470. <https://doi.org/10.1093/bioinformatics/bts172>.
- (25) Eyal, E.; Najmanovich, R.; Mcconkey, B. J.; Edelman, M.; Sobolev, V. Importance of Solvent Accessibility and Contact Surfaces in Modeling Side-Chain Conformations in Proteins. *J. Comput. Chem.* **2004**, *25* (5), 712–724. <https://doi.org/10.1002/jcc.10420>.
- (26) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447. <https://doi.org/10.1021/ct700301q>.
- (27) Hess, B.; van der Spoel, D.; Lindahl, E.; the GROMACS development team. GROMACS User Manual Version 4.6.1. **2013**.
- (28) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (29) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in

- Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, 25 (11), 1400–1415. <https://doi.org/10.1002/jcc.20065>.
- (30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, 79 (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (31) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, 26 (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.
- (32) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, 25 (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (33) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, 97 (40), 10269–10280. <https://doi.org/10.1021/j100142a004>.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; O. Yazyev, A.; Austin, J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
- (35) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* Basis Set for Atoms K through Zn. *J. Chem. Phys.* **1998**, 109 (4), 1223–1229. <https://doi.org/10.1063/1.476673>.
- (36) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Math. Proc. Cambridge Philos. Soc.* **1928**, 24 (1), 111–132. <https://doi.org/10.1017/S0305004100011920>.
- (37) Sanchez-Ruiz, J. M. Differential Scanning Calorimetry of Proteins. In *Proteins: Structure, Function and Engineering*; Biswas, B. B., Roy, S., Eds.; Plenum Press: New York, NY, 1995; pp 133–176. https://doi.org/10.1007/978-1-4899-1727-0_6.
- (38) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci.* **2018**, 115 (21), E4758–E4766. <https://doi.org/10.1073/pnas.1800690115>.
- (39) Perozzo, R.; Folkers, G.; Scapozza, L. Thermodynamics of Protein–Ligand Interactions: History, Presence, and Future Aspects. *J. Recept. Signal Transduct.* **2004**, 24 (1–2), 1–52. <https://doi.org/10.1081/RRS-120037896>.