Universidad Zaragoza
1542

Trabajo Fin de Máster

Estudio de las propensidades estructurales de la sequencia de alfa sinucleína

Study of the structure-propensities of alpha-synuclein sequence

Autor

Adrian López Borbón

Directores

Dr. Nunilo Cremades Casasín

Dr. Pierpaolo Bruscolini

FACULTAD DE CIENCIAS

2019

Acknowledgments

First of all I would like to thank my TFM directors, Dr. Pierpaolo Bruscolini and Dr. Nunilo Cremades because of their guidance and help.

I would like to thank also to Pablo Gracia and José D. Camino because of their continuous help and support and for everything they taught me.

To the Carolina Foundation and the University of Zaragoza for giving me this great opportunity for professional and personal growth.

To my family and Lisset who supported me from the distance.

Index

Abstract

The amyloid aggregation of alpha synuclein (αS), an intrinsically disordered protein, plays a key role in the etiology of Parkinson disease. The molecular mechanisms of αS amyloid aggregation are not well described at the moment and in this research, we move the first steps towards the characterization of two different mechanisms of αS aggregation as well as the of the phase transition between them using methanol (MetOH) as an inducing agent, with an approach that will combine the experimental and theoretical perspective with a residues-based theoretical modelling.

On the one hand the determination of the phase transition range between the two observed mechanisms was carried out by performing αS aggregation kinetics at different MetOH concentrations and carrying out structural analysis of the generated aggregates by Fourier Transform Infrared Spectroscopy (FTIR). On the other hand, the preference for one mechanism or the other was studied with pyrene excimer emission fluorescence spectroscopy and the coexistence of both types of aggregation mechanisms was observed in the transition phase. The results obtained from FTIR and emission fluorescence spectroscopy were in agreement and complementary to each other.

From the theoretical perspective an initial approach for the αS aggregation modelling was given by computing the accuracy of online secondary structure predictors in order to determine the importance of different αS amino acid regions in the preference for the type of amyloid aggregation mechanism.

Introduction

Amyloid formation

Amyloid aggregation is a process by which peptides and proteins misfold and self-assemble into highly ordered ß-sheet-rich aggregates. A large number of peptides and proteins have been reported to be able to self-assemble first into the form of oligomers and eventually into insoluble fibrillar aggregates, adopting a distinctive cross-ß structure (1, 2). This structure can be acquired by a polypeptide chain in an aggregated form disregarding its residue sequence or folding topology, either if fully folded or if intrinsically disordered (1,3), which has led to the proposal that the amyloid state is a generic protein conformation (1,2).

Amyloid aggregation mechanisms

The molecular mechanisms by which a protein undergoes amyloid aggregation are of remarkable interest and have been approached by experimental and theoretical methods using the chemical kinetics formalisms. There are important problems to be solved, such as, how do the aggregation and the acquisition of the cross-ß structure take place (4).

The transition of a protein to the amyloid state is influenced by the intrinsic protein features as well as environmental conditions. The overall process of amyloid fibrils formation follows a nucleation–polymerization model, where soluble species undergo a nucleation step that yields oligomeric species which are then able to grow through monomer addition generating protofilaments and eventually mature fibrils as can be represented in Figure 1 (4). In some cases, a structural conversion at the level of oligomers has been observed, leading to a nucleation-conversion-polymerization model (5).

Monomers　　　　　Oligomers　　　　　Fibrils
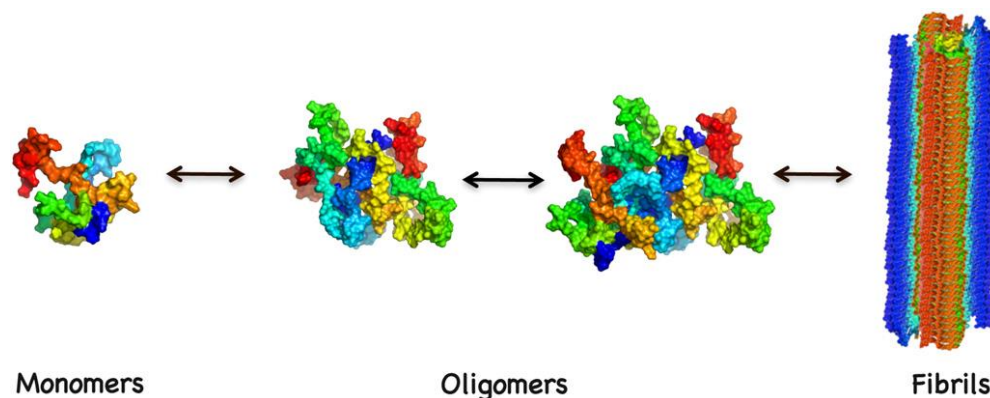
Figure 1. *Schematic representation of the process of amyloid formation (4).*

The typical sigmoidal profile of the aggregation process is indicative of the higher energy barrier to generate the first aggregate nuclei in comparison to the energy cost for aggregate growth.

Other processes than primary fibril nucleation and elongation have been reported recently as relevant such as fibril fragmentation and surface-catalyzed nucleation, which can largely influence the kinetics of fibril growth under many circumstances (4).

Alpha synuclein aggregation

Alpha-synuclein (αS) is an intrinsically disordered protein of 140 residues and a molecular weight of 14460 Da that is expressed abundantly in brain. Among its physiological functions it has been reported regulation of synaptic trafficking, homeostasis and neurotransmitter release (6,5). It is a major constituent of Lewy bodies and Lewy neurites, the hallmark of Parkinson´s disease, which is a neurodegenerative disorder that is related to the deposition of aggregated proteins (6).

The sequence of αS can be divided into three main regions: the N-terminal region (residues 1–60); the NAC (nonamyloid-ß component) region (residues 61–95), which is mainly hydrophobic; and the C-terminal region (residues 96–140), which is highly acidic, as it is shown in Figure 2 (4).

In the protein it can be found seven imperfect KTKEGV motifs between the N-terminal and the NAC regions which generate a hydrophobic periodicity similar to that found in the amphipathic lipid-binding α-helical domains of apolipoproteins (7). This feature results in the acquisition of an amphipathic α-helical structure in this region when binding to lipid membranes. In addition, αS is known to adopt ß-sheet structure upon the formation of amyloid fibrils.

The folding core of αS amyloid fibrils has been reported to be composed of five ß-strands located in residues between 30 and 100. Some analysis of the secondary structure of αS oligomers have revealed by Raman spectroscopy that in addition of ß-strands, some helical contribution is also observed (8). The latter information has led us to believe that a previous conformational change between alpha helix to beta sheet could be involved in the early stages of αS oligomers.
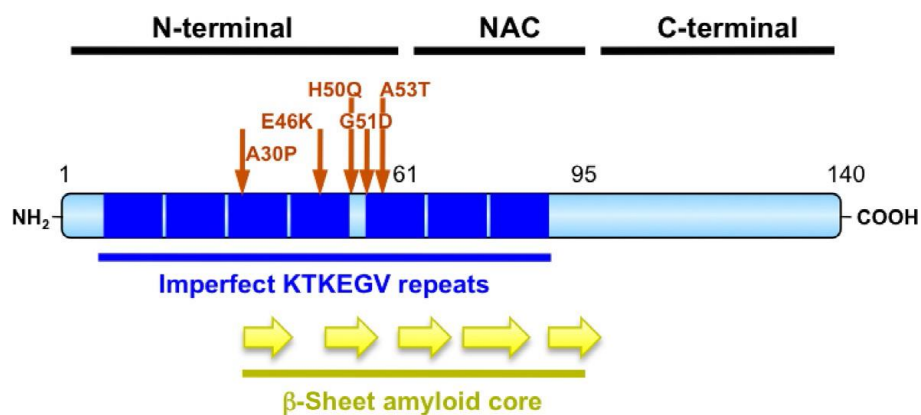
Figure 2. *Schematic representation of the sequence-dependent features of αS (4).*

αS aggregation can be reproduced in vitro and the fibrillary aggregates are morphologically and tinctorially indistinguishable from those isolated from PD patients (9). In vitro, αS aggregation has been reported to be initiated at hydrophobic/hydrophilic interfaces such as the air/water or lipid bilayer/water interfases (4, 7). In the presence of such interfaces, αS accumulates at the interfase given its amphipathic nature, and it is in this environment where it suffers heterogeneous nucleation (nucleation triggered by a particular surface). At those conditions, the process is strongly modulated by the pH and ionic strength of the solution, due to protective long range intramolecular interactions between the NAC region with the C-terminal region of the protein, which prevent the monomers from having aberrant interactions with the NAC region of other αS molecule in solution (7).

Structural features of αS fibrils

The core structure of αS fibrils generally includes residues 30–110 (10), although the exact position of the residues incorporated into the cross-β structure can vary in different fibril polymorphs.

Straight and twisted αS fibrils have been reported, where different fibril polymorphs can vary both in the specific molecular organization of the protofilaments and the arrangement and number of the protofilaments in the mature fibril (4).

Most experimental results are consistent with a model where αS monomeric units are stacked in a parallel arrangement forming the fibril protofilament (11, 12). Nevertheless, it has been documented the formation of fibrils arranged in an antiparallel geometry for other proteins prone to amyloid aggregation under certain experimental conditions (13, 14). In the case of αS, however, the antiparallel cross-β configuration has only been reported for a particular type of stable, kinetically trapped toxic oligomer (15).

αS amyloid aggregation in the presence of MetOH

As it was mentioned above the αS aggregation has been reported to take place through heterogeneous nucleation induced by surfaces. However, in the group where the presented work was performed, particular solution conditions for αS homogeneous nucleation (nucleation in the bulk of the solution) were identified, which can allow us to study the process of nucleation with more detail and to identify the factors and structural transitions in αS that are needed for the nucleation of the protein.

One of these conditions relies on the addition of low-to-moderate concentrations of alcohols such as methanol (MetOH).

Our research group has noticed that, in presence of MetOH αS aggregation is dramatically accelerated by homogeneous nucleation and depending on the MetOH concentration, differences in the kinetics profiles and the arrangement of monomeric units in the amyloid aggregates (parallel β-sheet vs antiparallel β-sheet aggregates) are observed. At low MetOH concentrations, αS fibrils are arranged in a parallel geometry, and a well-established lag phase is observed (Figure 3A). Conversely, above at certain MetOH concentration, αS fibrils are arranged in an antiparallel geometry and no lag phase is observed in the aggregation kinetics (Figure 3B). In addition, by FTIR structural characterization of αS aggregates obtained under the latter MetOH concentrations, a band is observed at about 1695 cm^{-1} (Figure 3D), which is a fingerprint for antiparallel β-sheet structures, whereas in the FTIR analysis of the aggregates obtained at low MetOH concentrations this band cannot be observed (Figure 3C).
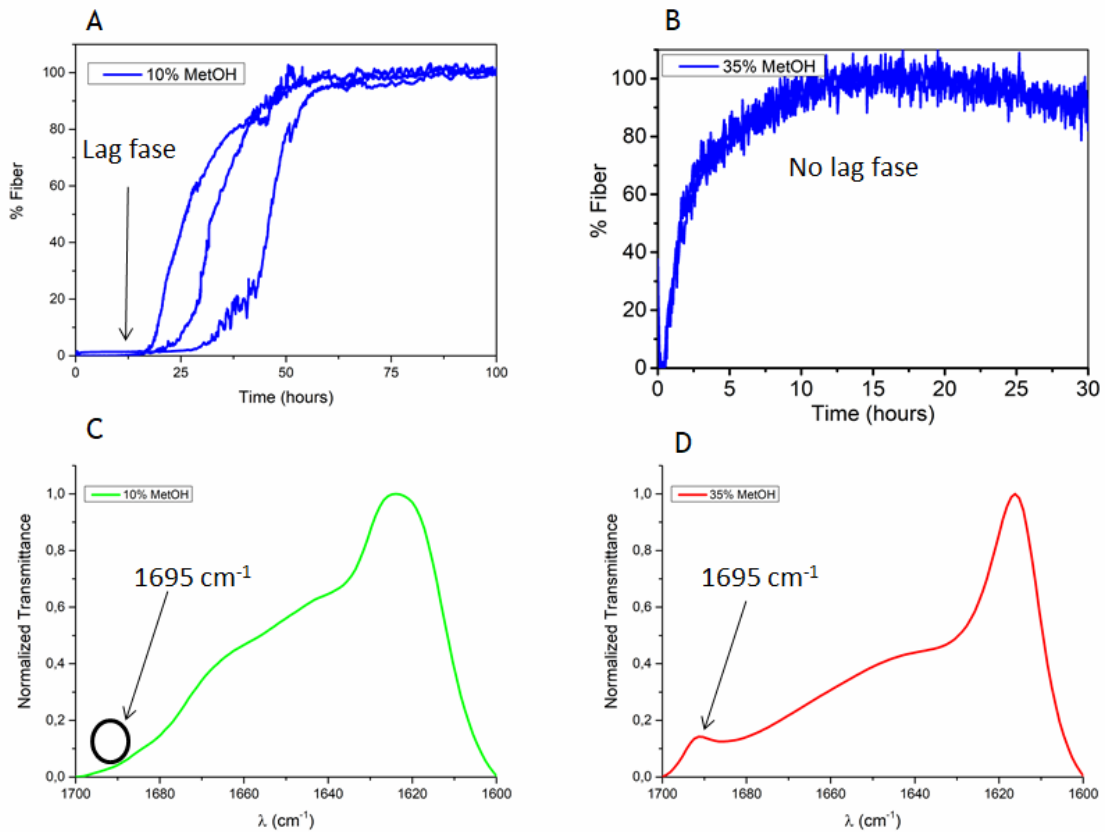
Figure 3. *aS parallel and antiparallel aggregation pathways at different MetOH concentrations. A: aS aggregates at MetOH 10% aggregation regime. B: aS aggregates at MetOH 35% aggregation regime. C: FTIR structural characterization of aS aggregates at MetOH 10% aggregation regime. D: FTIR structural characterization of aS aggregates at MetOH 35% aggregation regime. Data provided by José D. Camino.*

The present work aims at laying the experimental and theoretical basis for the detailed analysis of the early events of the two mechanisms of homogeneous nucleation identified in αS. Our goal in this work was to study the two different aggregation regimes presented in Figure 3 and to observe the phase transition between them. We intended to approach this by studying the structural characteristics of the alpha-synuclein aggregates, as well as their kinetics of aggregation, and also setting the basis of a theoretical modelling of the aggregation on a residue basis, by tuning the parameters of a simple statistical model, to reproduce the alpha-helical propensities of the αS sequence.

FTIR use in the oligomers and fibrils structural characterization: distinguishing monomers, parallel and antiparallel fibrils

FTIR occupies a key position in the analysis and comprehension of the complex aggregation mechanisms at the oligomer and/or fibril level, since amide I absorption band (1600–1700 cm^{-1}) spectra analysis is highly sensitive to changes in the H-bonding and secondary structure in proteins (16, 17).

It is well-established that, by FTIR, oligomers can be differentiated from fibrils simply on their spectral features, and therefore FTIR can be used as a tool for structurally follow a protein aggregation process. In addition, it is possible to tell apart different fibrils polymorphs by using this method. In particular, amyloid fibrils have a spectral signature

4

clustering between 1611 and 1630 cm^{-1} where fibrils in an antiparallel conformation exhibit an additional band at about 1695 cm^{-1} (17, 18).

Pyrene excimer fluorescence

Pyrene is a spatially sensitive probe which exhibits an ensemble of fluorescence emission peaks designated as bands I, II, III, IV and V at 375, 379, 385, 395 and 410 nm, respectively. This molecule has ideal properties as a probe for monitoring structural and dynamic changes during protein aggregation: a long fluorescent lifetime extending beyond 100 ns under suitable conditions, a relatively high extinction coefficient, excimer formation and solvent sensitive fluorescence spectra (19).

When two pyrene rings, one at the excited state and the other at the ground state are spatially proximal from each other (10 Å or closer) and with the proper orientation, an additional broad spectral emission band corresponding to the excimer species is originated at about 460 nm due to the interaction between those rings. The excimer/monomer (E/M) ratio is a measurement of the amount of excimer formed and its use has proved useful in the detection and characterization of early stages of αS aggregation as a distance indicator (20).

Since in αS parallel amyloid aggregates the monomeric units are in register in the ß-sheet amyloid core, the quantification of pyrene-labelled αS aggregates could be performed by measuring the E/M ratio.

Theoretical models used in αS aggregation

At the moment the models used to describe the amyloid aggregation mechanisms are mainly kinetics based models (6, 21, 22), which do not approach the amyloid aggregation process from a microscopic approach. Therefore, telling apart which regions of αS are involved in the amyloid aggregation and, specifically, in what stages of the aggregation they become relevant has been elusive until now.

A theoretical model which could provide insight in the aminoacids involved in the αS amyloid aggregation is a simple statistical-physics model based on the Wako-Sato-Muñoz-Eaton Model (WSME) (23-26). This is a highly simplified model of the protein folding process built on the premise that the latter is mainly determined by the structure of the native functional state, whose knowledge is assumed. Only native interactions are included in this model, but despite its simplicity it has been able to capture the main features of the kinetic behavior and folding pathways of specific molecules.

Zamparo and coworkers used the WSME method to develop an exactly solvable simplified statistical mechanical model for the thermodynamics of ß-amyloid aggregation (WSME-Agg) (27). Here the monomer concentration is explicitly taken into account as well as a nontrivial dependence on the microscopic degrees of freedom of the single peptide chain. However, in the original model only the case of an ideal helical structure, with homogeneous interactions, was considered to represent the protein's native state.

Since αS is intrinsically disordered, there is no deposited native structure in isolation[1] to use in the model, and since the latter is required to approach the αS aggregation through WSME-Agg, a reliable secondary structure predictor is required. Since as the αS is known to adopt alpha helix conformation upon binding to lipid membranes as described above, our first goal from the theoretical perspective was to find a reliable alpha helix predictor for the αS sequence, as an initial step for a theoretical understanding of the αS amyloid aggregation process.

---

1 The all-helical pdb structures 1XQ8 and 2KKW are micelle-bound, so one cannot be sure of the extent to which the interaction with the micelle interferes in stabilizing certain helical structures, on top of the intrinsic propensities.

Hypothesis:
We believe that using theoretical methods to integrate experimental information about αS stability and kinetics, we will be able to predict which are the key residues involved in αS misfolding during the early nucleation stages and the molecular mechanisms that govern this process, as a first step to develop a model that is realistic and detailed enough, to study the microscopic determinants, at the residue level, of the onset of aggregation, rationalizing the experimental findings on the structural properties and kinetics of amyloid formation.

General Goal:
Study by experimental and theoretical methods the key physical and chemical factors that define the stability and homogeneous nucleation in αS.

Specific Goals:
1 Determine the phase transition range between the formation of parallel and antiparallel β-sheet aggregates.
    1.1 Expression and purification of αS wild type (WT).
    1.2 Aggregation kinetics of αS WT with different MetOH concentrations.
    1.3 Structural characterization by Fourier Transform Infrared spectroscopy (FTIR) of the aggregates generated in the transition range.
2 Determine the fraction of parallel and antiparallel aggregates along the transition between the aggregation regimes.
    2.1 Fluorescence labelling of the mutant proteins αS-K6C, αS-Q24C, αS-A56C, αS-A85C and αS-A140C.
    2.2 Determine the most suitable labelled mutant protein to be used in the fibrils characterization by pyrene fluorescence spectroscopy.
    2.3 Determine the fraction of parallel and antiparallel aggregates under relevant MetOH concentration conditions.
3 Predict the alpha-helix propensity in the different regions of the sequence.
    3.1 Search on online alpha-helix probability predictors and determine the most accurate.

Materials and Methods

Expression of αS
Solutions

PBS buffer pH= 7.4: Was prepared by adding 62mg NaH_2PO_4 (Scharlau), 208mg Na_2HPO_4 (Scharlau) and 1,752g of NaCl (Scharlau) to a final volume of 200 ml using ultrapure type I H_2O (Mili-Q).

Lysis buffer: It was prepared by adding 2mL of stock solution Tris 1M (Scharlau) pH=7.7 to 400 µL of stock solution EDTA 0.5M (Scharlau) and 5 protease inhibitor cocktail tablets (Roche Diagnostics GmbH) to a final volume of 200mL using ultrapure type I H_2O (Mili-Q).

Anion exchange buffer A: 25mM Tris, pH=7.7.

Anion exchange buffer B: Tris 25mM, NaCl 1.5M pH=7.7.

Transformation
25 µl of BL21 Gold (DE3) competent cells (Agilent technologies) and 1 µl of plasmid were added to a 1.5 ml vial and were placed on ice for 30 min.

The plasmid o pT7-7 WT αS (Addgene #36046) was used, which codes for human αS, contains an ampicillin resistance gene and its expression can be induced by isopropyl ß-D-1-thiogalactopyranoside (IPTG).

A subsequent heat shock was given to the sample at 42 ºC for 20 sec using a metallic block thermostat (Tembloc, PSelecta). The sample was placed on ice for 2 minutes afterwards. 200 µl of LB Broth Lennox 20 g/l (Formedium) solution without antibiotic were added and the sample was incubated at 37 ºC and at 220 rpm shaking for 1 hour (stove MaxQ 400, Barnstead lab-line). 200 µl of sample were added to amplification plate and incubated at 37 ºC without shaking overnight (O/N).

Pre-culture
In a 250 ml Erlenmeyer sterile flask 50 ml of LB culture medium and 50 µl of 100 mg/ml ampicillin were added and grown at 37 ºC and 230 rpm shaking O/N.

Scaling up and Induction
Three Erlenmeyer flasks containing 2L of LB medium each were autoclaved and afterwards 1ml of ampicillin (100 mg/ml) and 5ml of starter culture per 1L of LB medium were added in sterile conditions.

The samples were incubated at 37˚C and 220 rpm shaking (Certomat BS-1, Sartorius stedim biotech) until an optical density of about 0.65 at λ= 600 nm was reached ($OD_{600}≈0.65$) (Spectrophotometer Ultrospec 6300 Pro, Amersham Bioscience). Afterwards 1ml of IPTG 1M per 1L of culture medium was added and incubated 4 hours at 37˚C at 220 rpm shaking, in order to induce αS expression.

The cell culture was centrifuged (Avanti J-E Centrifuge, Beckman Coulter) using 500 mL tubes (Beckman Coulter) in a JA10 rotor for 10 min at 4 ºC and 9700 rpm, throwing the supernatant away. Pellets were resuspended in PBS buffer, and centrifuged again this time using JA20 rotor for 30 min at 4 ºC and 13000 rpm. The pellet was stored at -80 ºC.

Purification of αS

The pellet was resuspended in 20 ml of lysis buffer per 1L of culture. The solution was sonicated (Ultrasonic Processor, Sonics VibraCell) in ice for 5 cycles at a power of 80 for 1 minute with a 1 minute pause to break the cells. A centrifugation step was carried out using a JA-20 rotor for 30 minutes at 4 °C, at 13000 rpm. The pellet was thrown away.

Protein precipitation by heating

The precipitation of heat-sensitive proteins was carried out by heating for 20-25 minutes at 80-95 °C (Frigiterm, PSelecta), since alpha synuclein is an intrinsically disordered protein, it remained soluble. The sample was centrifuged as before using a JA-20 rotor for 30 minutes at 4 °C, at 13000 rpm. The pellet was thrown away.

DNA precipitation

Afterwards DNA was precipitated by adding 10 mg of streptomycin sulphate (Carbosynth) per 1 ml of supernatant and was stirred for 15 minutes at 4 °C. The sample was centrifuged as before using a JA-20 rotor for 30 minutes at 4 °C, at 13000 rpm. The pellet was thrown away. The DNA precipitation and subsequent centrifugation steps were repeated since DNA is the main contaminant of *aS*.

aS precipitation

aS was precipitated by adding slowly 361 mg of ammonic sulphate (Roth) per 1 ml of supernatant and stirring for 30 minutes at 4 °C. The sample was centrifuged as before using a JA-20 rotor for 30 minutes at 4 °C, at 13000 rpm. The supernatant was thrown away and the pellet was resuspended in Tris 25mM pH= 7.7, 6-7 mL per 1L culture volume.

Dialysis

A 3500 Da membrane (Spectra/Por) and 4L 25mM Tris, pH=7.7 buffer were used in dialysis at 4 °C under stirring. The buffer was changed twice after three hours.

Anionic Exchange Fast Protein Liquid Chromatography

The solutions were filtered and degassed using a 0.22 μm filter (MicroSep nitrocelullose, Life Sciences); whereas the protein sample was filtered using a 0.45 μm filter (MicroSep nitrocelullose, Life Sciences). Through the anionic exchange column (GE Healthcare HiPrep QFF 16/10) 5 column volumes (C.V.) of ultrapure type I (Mili-Q) H2O were passed to remove the ethanol in which the column was embedded, followed by 5 C.V. of buffer A to equilibrate the column. The sample was injected at 2mL/min in order to allow the protein to bind the column matrix and 3 C.V. of buffer A were passed at 5mL/min to remove unspecifically bound substrates. Afterwards 5 C.V. of buffer B 7.5% were passed to remove impurities. Then 20% of buffer B was passed, where alpha synuclein was eluted and fractions every 3.5 mL were collected. In order to remove DNA from the column, 5 C.V. of 100% buffer B were passed. Finally the column was cleaned and equilibrated with MQ-water and 20% ethanol respectively.

Gel Filtration Fast Protein Liquid Chromatography

Size exclusion chromatography was used to separate the protein oligomers, monomers and fragments. The solutions were filtered and degassed using a 0.22 μm filter (MicroSep nitrocelullose, Life Sciences); whereas the protein sample was filtered using a 0.45 μm filter (MicroSep nitrocelullose, Life Sciences). It was passed 1 C.V. of ultrapure type I (Mili-Q) H2O through the column (HiLoad 26/600 Superdex 75pg) in order to remove the ethanol

9

in which it was embedded. Afterwards the column was equilibrated by passing 1 C.V. of PBS pH= 7.4 buffer without exceeding a flow rate of 2ml/min and a pressure of 0.3MPa. The protein had been previously concentrated to 0.5mL every 1L of culture using centrifugal concentrators (Merck Millipore) of 10kDa. The protein sample was loaded and afterwards eluted with PBS pH=7.4 at 2mL/min. When the dead volume of the column (100 mL) eluted the fractions were collected every 3.5 mL. The collected protein was concentrated using centrifugal concentrators (Merck Millipore) and quantified using a nanodrop spectrophotometer (NanoValue Plus, Biochrom) by measuring absorbance at λ=275 nm and computing the concentration using the Lambert-Beer Law with ε275nm= 5600 M^{-1} cm^{-1}. The protein was stored at -80 ^{o}C.

Aggregation kinetics of αS WT with different MetOH concentrations

Thioflavin use in amyloid formation kinetics
Thioflavin T (ThT) is a highly suitable amyloid dye as it yields a high fluorescence signal at about 482 nm upon excitation at 450 nm when bound to amyloid. The mechanism of fluorescence enhancement upon binding to amyloid has been attributed to the rotational immobilization of the central C–C bond connecting the benzothiazole and aniline rings (28).

The analysis of experimental kinetic data for the fibril formation process using ThT has significantly enhanced the understanding of the mechanisms of the fibril formation process. However, characterization of the nucleation events themselves is challenging as a consequence of the elongation rates of the transient intermediate species and their high heterogeneity (4), and in this particular case ThT is unable to bind small oligomers due to size requirements for its binding (29).

αS WT aggregation kinetics
The αS WT aggregation kinetics was analysed by fluorescence in a multiplate fluorimeter (FLUOstar® Omega, BMG LABTECH). The experiment was carried out using in each well 100 μM αS WT, 50 μM ThT (Carbosynth) and 0.01% in volume NaN3 (Scharlau) in PBS buffer at pH=7.4. The αS aggregation was performed using MetOH volume percentages of 0%; 5%; 10%; 15%; 20% and 25%. Measurements were carried out using an excitation wavelength of 440 nm and emission wavelength of 480 nm with a 10 nm window.

FTIR characterization of αS WT fibrils
The samples from the wells were collected and centrifuged in an ultracentrifuge (Optima MAX-XP Ultracentrifuge, Beckman Coulter) and the solvent was removed and changed by PBS buffer prepared in deuterated water. The ultracentrifugation followed by buffer change process was repeated twice. The measurement was carried out in the 1000-4000 cm^{-1} range. Afterwards the mid infrared region, specifically from 1500-1900 cm^{-1} was selected and the buffer was subtracted. This experiment as well as data processing was performed by the PhD student José D. Camino.

Fluorescence labelling of the mutant proteins
The mutant proteins (αS-K6C, αS-Q24C, αS-A56C, αS-A85C and αS-A140C) were buffer-exchanged into Tris 25mM NaCl 150mM supplemented with tris(2-carboxyethyl)phosphine (TCEP) (Carbosynth) at a concentration 5 times higher than the concentration of cysteine in the proteins samples using PD10 (GE Healthcare) desalting columns. This eliminated the dithiothreitol (DTT) from the sample allowing further maleimide-thiol chemistry labelling

of the protein. The proteins were incubated at 4ºC with pyrene-5-maleimide (Santa Cruz Biotechnology) in a concentration 5 times higher than the concentration of cysteine in proteins, under gentle mixing overnight. The probe excess was quenched using by adding DTT (Carbosynth) at a concentration 5 times higher than the pyrene maleimide concentration, for half an hour. The free, unreacted probe was separated from the mixture with a PD10 column using buffer Tris 25mM NaCl 150mM pH=7.4. The obtained proteins were concentrated using 10 kDa cutoff centrifugal concentrators and quantified by measuring absorption in a nanodrop instrument at λ=343 nm, using an extinction coefficient of 34700 L*mol^{-1}*cm^{-1} (31).

The reaction products were analysed by Mass Spectrometry (MS) in order to confirm the attachment of the probe to the protein. This was carried out at CIBA as external service in a MALDI-TOF/TOF Mass Spectrometer.

Screening for the determination of the most suitable pyrene-labelled mutant protein
The pyrene-labelled mutant proteins (αS-K6C-Pyr, αS-Q24C-Pyr, αS-A56C-Pyr were αS-A85C-Pyr) were mixed with in a 1:4 ratio for a total protein concentration of 50 μM, in Tris 25mM, NaCl 150 mM, 0.01% NaN3 and TCEP 200 μM. Aggregation was induced by adding MetOH volume percentages of 10% and 35% to the samples in a 96-well-plate.Three replicates were measured for each condition and an excitation wavelength of 343 nm and emission wavelengths of 375nm; 385nm and 470nm, using a 5nm window in a multiplate fluorescence reader (CLARIOstar).

At the final point of the aggregation the fluorescence intensity emission spectra of all the aggregates was measured (Cary Eclipse Fluorescence Spectrophotometer) using an excitation wavelength of 343 nm and cuvettes (Hellma Fluorescence Cuvettes). The E/M ratio of the retrieved aggregates was assessed.

Determination of the fraction of parallel and antiparallel aggregates with relevant MetOH concentration conditions

aS-A85C pyrene-labelled aggregation under relevant MetOH concentration conditions
The experimental procedure was the same as the one used for the previous screening and, by inducing the aggregation this time with MetOH concentrations of 10%; 15%; 16.5%; 18%; 20%; 22.5%; 25% and 35%.

The E/M ratio (FI (470 nm)/ FI (376 nm)) was used for the quantification antiparallel and parallel fibrils on each aggregation condition.

Selection of a reliable secondary structure predictor
As a disordered protein, alpha-synuclein does not present a stable native structure in isolation, so that we cannot rely on an experimentally-determined secondary-structure tendency. In order to have a reference secondary-structure propensities, against which to parameterize the theoretical models, we have to rely on secondary structure predictions by on-line servers. However, different servers usually give somewhat different predictions, so that it is important to assess and compare the quality of their predictions, in order to choose the best one . To test their performance, the online servers NetSurfp2.0[2] Jpred[3],

---

2 http://www.cbs.dtu.dk/services/NetSurfP-2.0/

Agadir[4] and Sopma[5] were used to predict the helix propensity of a total of 3370 residues of 26 proteins, whose PDB files were known: 1a5e, 1a23, 1ayf, 1bni, 1bpi, 1bta, 1c9o, 1chk, 1cun, 1cyo, 1fkj, 1ftg, 1gd1, 1h7m, 1hfy, 1otr, 1pga, 1rhg, 1rx4, 1sce, 1stn, 2ci2, 2rn2, 2trx, 3ssi and 451c. The results were compared to the ones reported by DSSP on the PDB files. The proteins were chosen from the thermodynamic database ProTherm according to the criteria of being all-alpha proteins and covering a wide range of thermal stability. The data were manipulated using mainly bash scripts in Linux (Appendix 3) in order to minimize manipulation errors, yielding, for each predictor, a single two columns file, containing, respectively, fthe residue one-letter code and the predicted alpha helix propensity fraction . Analogously, a two-columns file was built with teh one-letter codes and the DSSP output from the PDB files.

The predictions by different servers were quantitatively analyzed resorting to a Receiving Operating Characteristic (ROC) curve (calculated using the octave script in Appendix 1, where the true positives rate or sensitivity (TPR) was plotted in the y axis against false positive rate or 1-specificity (FPR) in the x axis, for every selected threshold value:

$$TPR = TP/(TP + FN)$$

$$FPR = FP/(FP + TN)$$

Where:

TP= True Positives; FP= False Positives; TN= True Negatives and FN= False Negatives. These values were determined by comparing the online predictor's values of alpha helix propensity to the value given by the DSSP output for each residue at a given threshold, in the following way: residues reported by DSSP as helical are considered as Positive instances; then if the probability score, predicted by the online server for those residues, is above a given threshold, they are considered as True Positives, and False Negatives otherwise.

Thus, at any value of the threshold, a certain value for the FPR and TPR is determined, yielding a point on the graphics. By varying the threshold, it is possible to span all the range of possibilities: from all residues predicted as positives, to all residues predicted as negative, and the full ROC curve is recovered.

The area under the curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. Since the highest value for this area is 1, the closer the value of area under the curve is to 1 the more accurate is the predictor.

In a ROC curve a perfect prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC plane, representing 100% sensitivity (no false negatives) and

---

3 http://www.compbio.dundee.ac.uk/jpred4/index_up.html

4 http://agadir.crg.es/

5 https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html

100% specificity (no false positives). The (0,1) point is also called a perfect classification. In general, the threshold value that yields a point with the highest height from the diagonal in the ROC curve would be the optimal threshold for a given predictor. Such criterion for determining the best probability threshold to identify helical residues corresponds to maximizing

the Youden's index is defined as:

$$J = sensitivity + specificity - 1 = TPR - FPR$$

This statistic test captures the performance of the predictions taking values form 1 to -1 and its maximum value may be used as a criterion for selecting the optimum cut-off point in a diagnostic test.

The accuracy index is another indicator that reflects the probability that the test has made a correct diagnosis. To calculate it, we construct a quotient placing in the numerator all possible true values (positives and negatives) and in the denominator all possible outcomes:

$$accuracy\,index = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

The threshold value that yields a point with the highest value of accuracy index would be the optimal threshold for a given predictor.

In the octave script there was also included a comparison between the DSSP file and the predictors' files in order to check if there was any mismatching in the residues sequence.

Results and discussion

Expression and purification of αS WT

The protein was over-expressed in BL21 Gold (DE3) *E.coli* cells and then purified using several precipitation and chromatographic steps as explained in Materials and Methods Section. After the different precipitation steps to remove the nucleic acids by streptomycin precipitation as well as most of the non-desired proteins by heat denaturation and salting out aggregation, the protein solution was eluted to an anion exchange column to remove the remaining DNA.

The chromatogram obtained (Figure 4) showed αS starting eluting at a NaCl concentration of 0.3 M (20% of Buffer B), whereas a large DNA peak is observed when 100% buffer B is used as DNA is more strongly retained in the matrix.
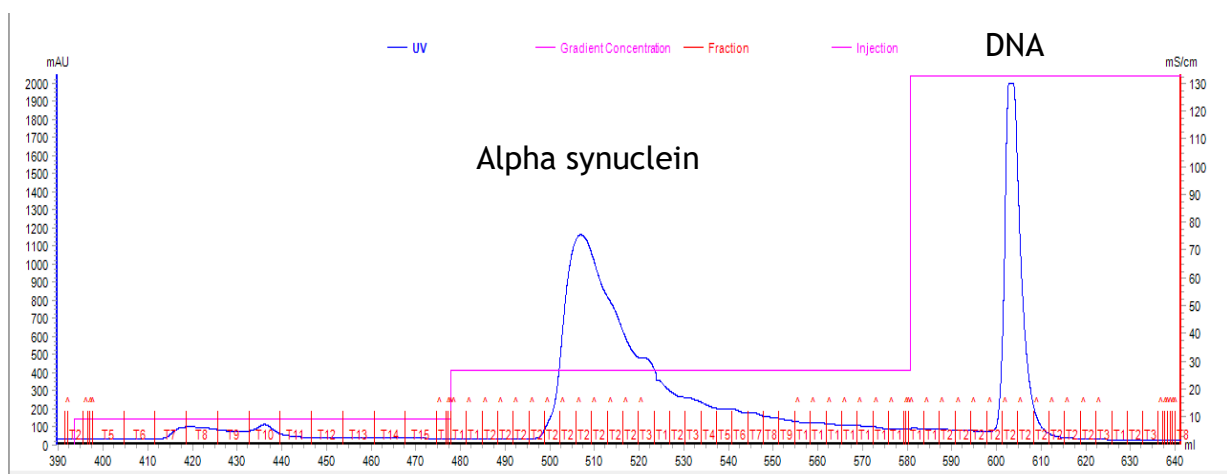

Figure 4. *Anionic exchange chromatogram*.

After this chromatographic step, the sample had been purified in αS, athough it required a further step in order to remove possible αS aggregates and fragments. In order to achieve this, the sample was loaded into a gel filtration column. As can be observed in Figure 5, the monomeric fraction of αS WT could be isolated by collecting fractions 9-26.

The purity of the protein obtained was assessed by polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 6) where it was observed that the collected protein exhibited a single migration band corresponding with about 14kDa which coincides with the expected molecular weight of αS WT (14460 Da).
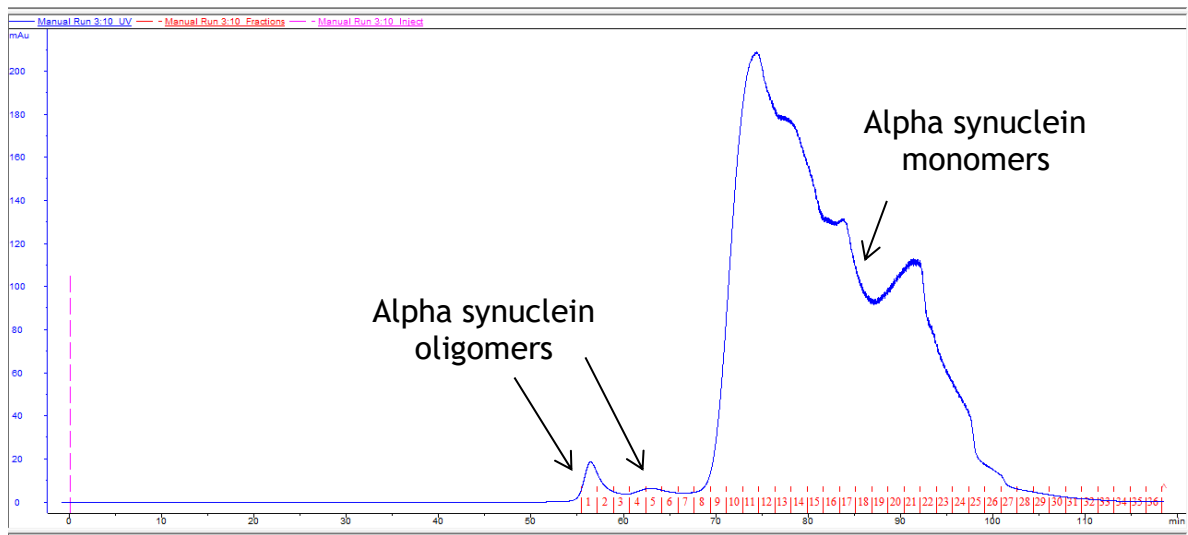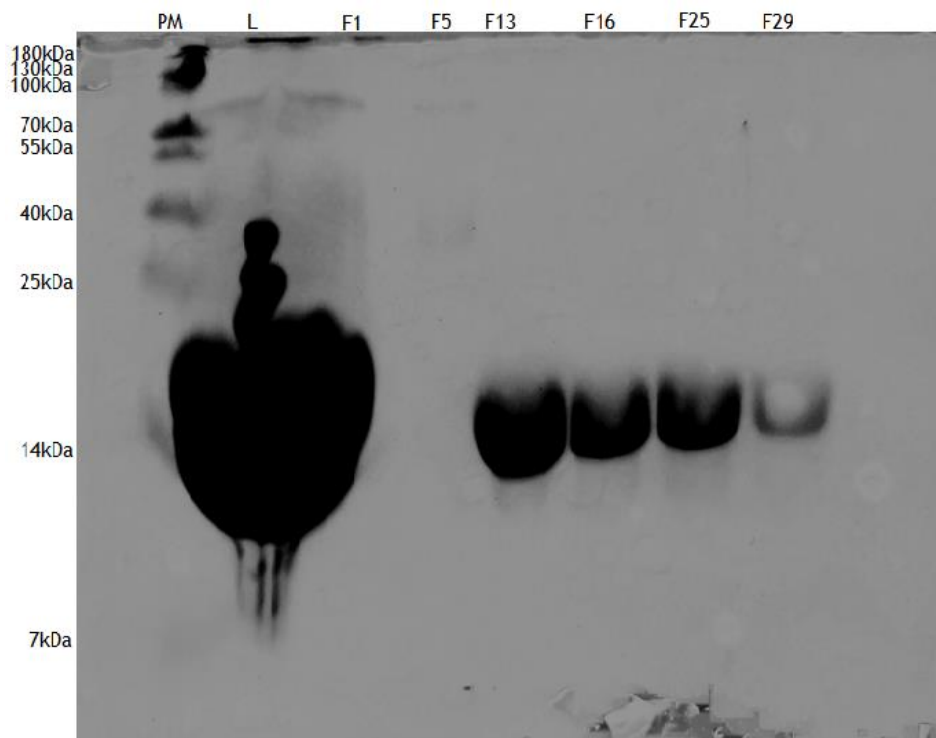
14

Figure 5. *Gel filtration chromatogram.*

Figure 6. *SDS-PAGE of the aS WT purification. The samples contained are PM: Prestained molecular marker (PageRuler Prestained Protein Ladder, ThermoFisher). L: Load of the aS WT to the gel filtration column. F1: Fraction 1 collected form the gel filtration chromatography. F5: Fraction 5 collected form the gel filtration chromatography. F13: Fraction 13 collected form the gel filtration chromatography. F16: Fraction 16 collected form the gel filtration chromatography. F25: Fraction 25 collected form the gel filtration chromatography. F29: Fraction 29 collected form the gel filtration chromatography.*

The αS WT purification process yielded about 24 mL at 309.5 μM from 6 L of cell culture and it was stored in aliquots at -80 °C.

Aggregation kinetics of αS WT at different MetOH concentrations

As it was previously mentioned, αS exhibits two different aggregation mechanisms depending on the MetOH concentration in the sample and therefore there are two different regimes where either parallel or antiparallel ß-sheet aggregates formation is favored. With the aim of studying these two mechanisms and especially the confluence of the both regimes at intermediate MetOH concentrations, on the one hand aggregation kinetics experiments were performed in the regime of low MetOH concentrations in order to monitor the influence of MetOH in the protein nucleation and on the other hand a deeper analysis on the transition phase was carried out.

Aggregation of αS WT was performed using ThT fluorescence emission under different MetOH concentrations (0%, 5%, 10%, 15% and 20%) in order to explore the effect caused in the kinetics profiles by the variation of dehydration conditions.

In the kinetics experiments (Figure 7A) it could be observed a profile change upon increasing MetOH concentrations. The lag time dropped when the MetOH concentration was incremented which is indicative of an acceleration of the overall rate of aggregation and particularly the rate of primary nucleation. This can be more clearly evidenced in the correlation of the apparent lag time with MetOH concentration in Table 1. As a consequence of this behavior the exponential growth phase starts earlier for higher MetOH concentrations and the plateau is reached faster also. The strong differences in the rate of aggregation of the protein by small changes in MetOH concentration reinforces the assumption that amyloid aggregation is heavily influenced by environmental conditions and the latter can drive the formation and coexistence of polymorphs. This observation also provides evidence that the different amyloid aggregation pathways are competing and one could be favored over the other only by changing the reaction conditions.
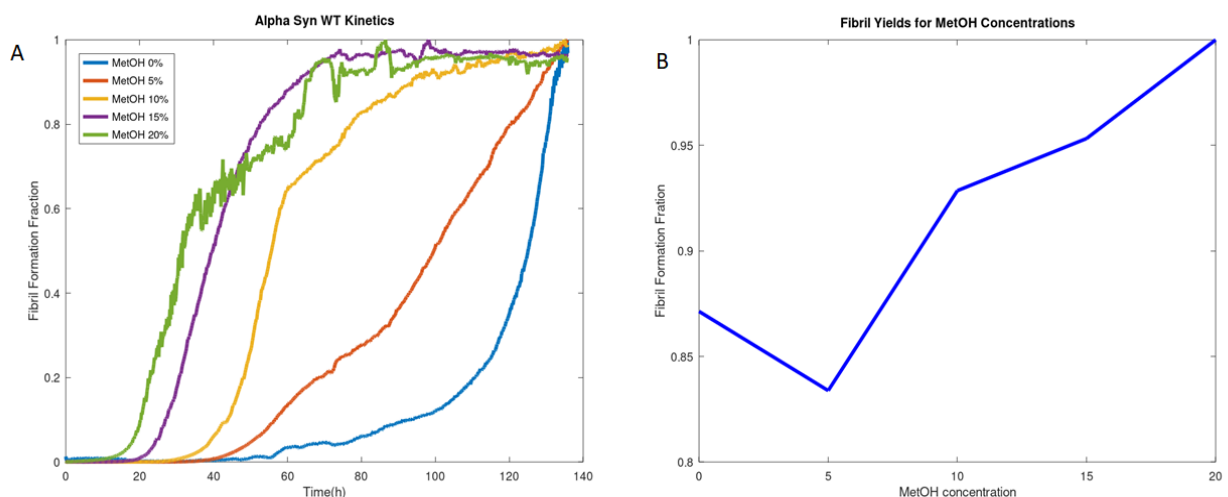

F**igure 7.** *A: αS fibril formation fraction time for different MetOH concentrations, monitored by ThT emission fluorescence. **B**: αS WT fibril formation fraction dependence with MetOH concentration, measured by SDS-PAGE yield analysis.*

Table 1. *Lag time for the aS WT aggregation under different MetOH conditions*

	MetOH 0%	MetOH 5%	MetOH 10%	MetOH 15%	MetOH 20%
Lag time(h)	92.6	56.5	43.7	26.8	20.3

The αS WT fibrils' yield was computed through SDS-PAGE by using an intern standard of αS WT at 100 μM which was the protein concentration used in the kinetics experiments. The samples were centrifuged (Optima MAX-XP Ultracentrifuge, Beckman Coulter) and the supernatant was loaded into SDS-PAGE gel (Figure 8) in order to estimate the monomer concentration remaining in solution. By image processing the band areas were retrieved for each sample and by comparing them to the area provided by the standard.
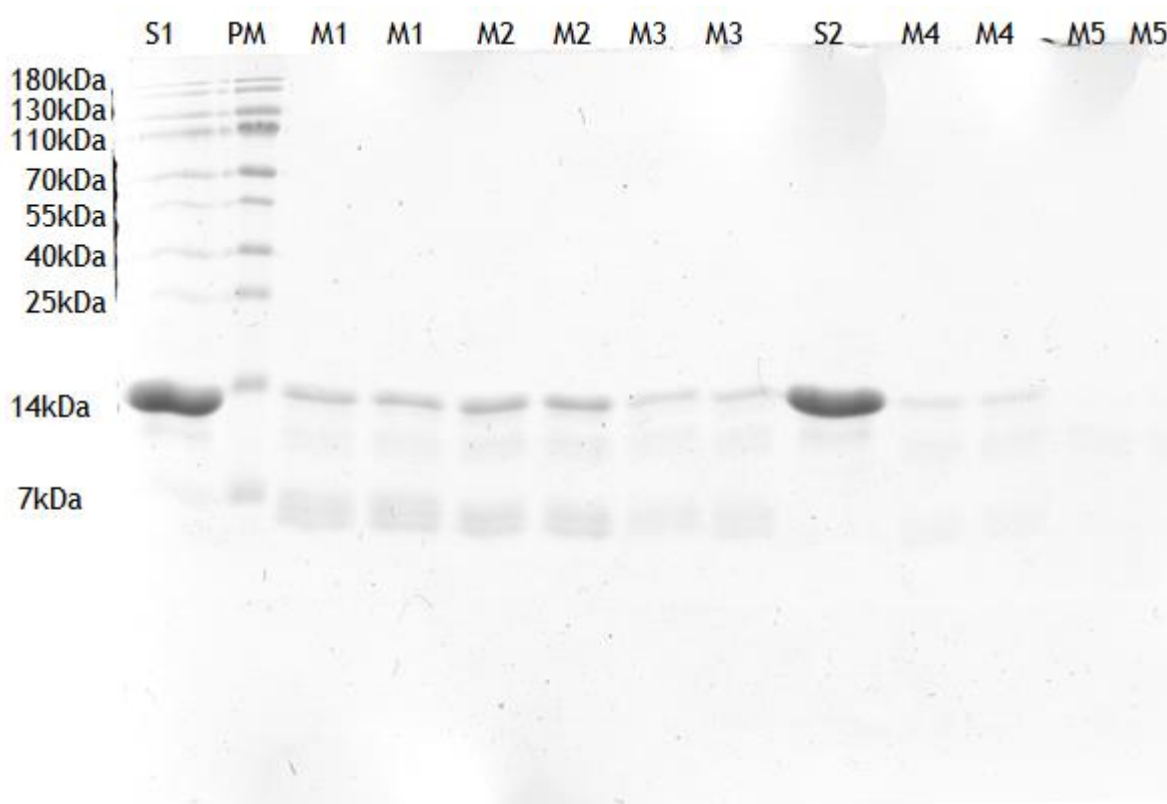

Figure 8. *SDS-PAGE for aS WT fibrils yield quantification. S1 and S2: aS WT standard 100 μM. PM: Prestained molecular marker (PageRuler Prestained Protein Ladder, ThermoFisher). M1: Replics of MetOH 0% condition. M2: Replics of MetOH 5% condition, M3: Replics of MetOH 10% condition, M4: Replics of MetOH 15% condition, M5: Replics of MetOH 20% condition.*

In spite that S1 and S2 are two replics of the αS WT standard, there were observed some bands in S1 well due to a manipulation error where there was contamination with prestained molecular marker, for that reason only S2 was used as reference in the relative quantification. By the fibril quantification it could be observed that not only the reaction

aggregation rate was increased upon increasing MetOH concentration but also the reaction yield was improved (Figure 7A, Figure 7B).

The dramatic acceleration of the aggregation process in αS by moderate concentrations of MetOH could be associated with changes in the environment polarity, since it has been reported conformational changes in αS due to contact with nonpolar interphases, inducing a partial adoption of alpha helix configuration. It is possible that a previous transient alpha helix configuration could make easier the adoption of a ß sheet configuration characteristic of the amyloid aggregates cores. The working hypothesis is, that the entropic cost for a transition from a partially folded molecule to a highly organized structure, as the amyloid core, is lower than the corresponding one for a transition from a completely disordered protein as it is the case of αS which would have to lose more freedom degrees.

FTIR characterization of αS WT fibrils

The resulting fibrils from the aggregation of αS WT at MetOH concentrations of 5%, 10%, 15%, 20%, 25%, 30%, 35% and 40% were analyzed by FTIR for structural characterization. The FTIR analysis were carried out by the researcher and PhD student José D. Camino. The FTIR analysis was performed in the amide I region (1600-1700 cm^{-1}) where the C=O stretching mode of the amide functional group plays a predominant role (about 80%) and the remaining contribution (20%) arises mainly from C-N stretching.
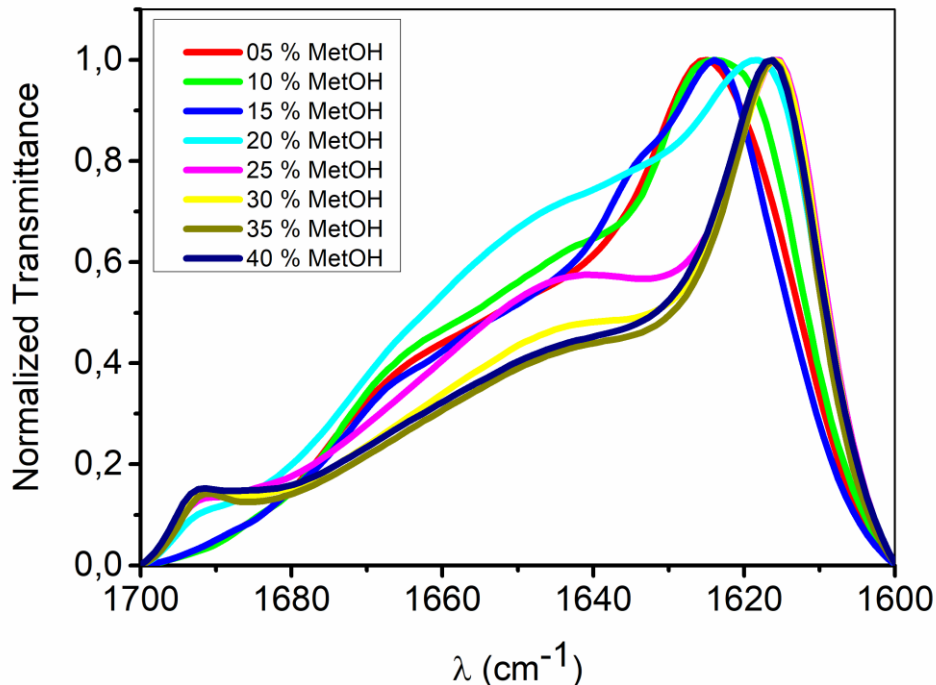

Figure 9. *FTIR structural analysis of αS WT fibrils formed at different MetOH concentrations.*

As can be observed in Figure 9, a change in aggregation regime is observed under MetOH titration conditions. This is evidenced by the presence of a band maximum at about 1625 cm^{-1} which is characteristic of the amyloid cross-ß structure and absence of a band located at about 1695 cm^{-1}, features characteristic of a parallel ß-sheet organization that can be observed at 5%, 10% and 15% MetOH concentrations. The ß index ratio (defined as the 1695 cm-1/1625 cm-1 intensity ratio), has been shown to be proportional to the percentage of antiparallel arrangement of ß-strands in a ß-sheet (17). At higher MetOH concentrations (20%, 25%, 30%, 35% and 40%) the typical signature of antiparallel ß-sheet organization index is observed in the aggregates: the presence of both the 1625 cm^{-1} and the 1695 cm^{-1} bands. Indeed, the 1625 cm^{-1} band in these aggregates is slightly shifted to lower wavenumbers (up to 1616 cm^{-1}), likely due to the formation of shorter intermolecular H bonds, as a consequence of fewer constraints on the peptide backbone (16).

The ß index ratio was plotted against MetOH concentration (Figure 10) as the intensity ratio between the two spectral contributions of ß-structure (I_{1695}/I_{1625}) is proportional to the percentage of anti-parallel vs. parallel organization of the ß-strands.
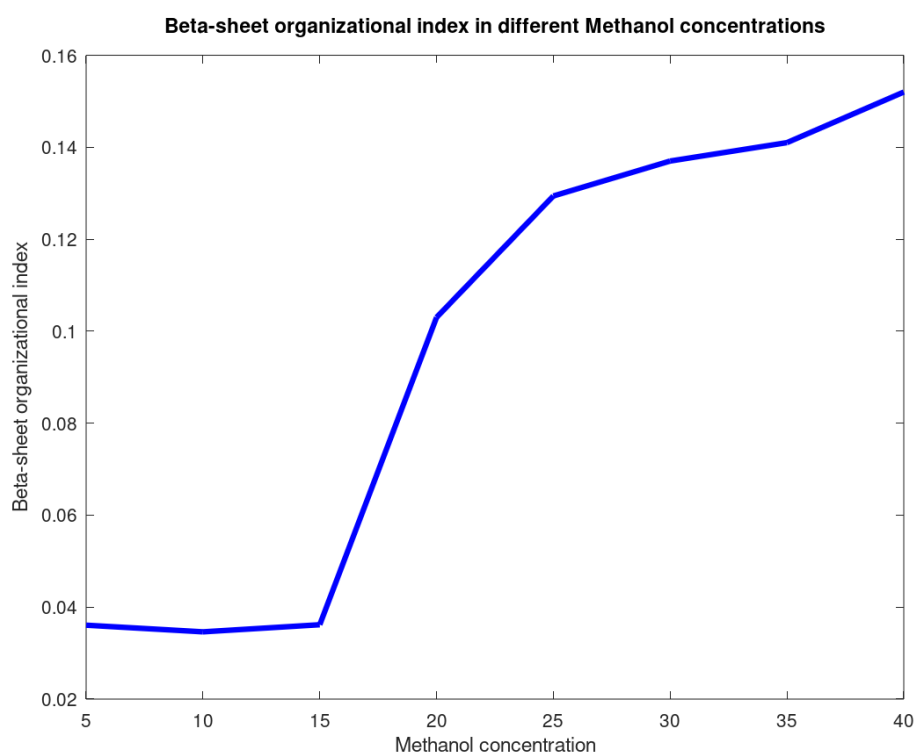

Figure 10. *ß-sheet organizational index for different MetOH concentration values.*

For MetOH concentrations lower than 15% the αS aggregates formed shoed typical in register parallel ß-sheet conformation, while at concentrations above 25%, the aggregates formed have primarily an antiparallel ß-sheet conformation. At concentrations between 15% and 25% of MetOH, the system is in an intermediate regime. In order to study this intermediate regime in more detail, a quantification of the percentage of parallel and antiparallel ß-sheet structures as a function of MetOH concentrations was required. To

achieve this goal, it was used the particular properties of pyrene excimer to estimate the percentage of parallel ß-sheet aggregates when the aggregation reaction was completed at different concentrations of MetOH.

Fluorescence labelling of the mutant proteins
In order to introduce a pyrene molecule in the protein that could report on the structure of the amyloid aggregates formed at different MetOH concentrations, cysteine (Cys) mutants of αS (which naturally lacks this type of aminoacid in its WT sequence) were generated and purified in the laboratory. In this work, a selection of αS-K6C, αS-Q24C, αS-A56C, αS-A85C and αS-A140C were used in labelling reactions in order to introduce a pyrene molecule upon reaction of the side chain of the Cys residue (concretely the SH group) with the maleimide group of the pyrene-maleimide compound that was purchased from commercial vendors (see Methods section). The yield of production of pyrene-labelled αS in each αS variant was later analyzed by mass spectrometry (Appendix 1). Interestingly, despite αS being an intrinsically disordered protein, and then expected similar solvent exposition of the different positions used to introduce the pyrene, the reaction yields for different positions showed remarkable differences , which could be given by the different chemical environment surrounding each Cys position along the amino acid chain. The αS-A140C variant was discarded in this step due to the very poor labelling reaction yield.

Screening for the determination of the most suitable pyrene-labelled mutant protein
Aggregation of the mutant pyrene-labelled proteins was carried out under MetOH 10% and 35% in order to assess the pyrene labelled proteins capability to provide information on the resulting parallel and antiparallel nature of the amyloid aggregates generated at the two reference conditions. As can be observed in Figure 11A, the fibrils obtained under parallel aggregation regime (MetOH 10%) exhibited a fluorescence emission band with a maximum at λ= 470 nm corresponding with the excimer formation due to the interaction between pyrene molecules.
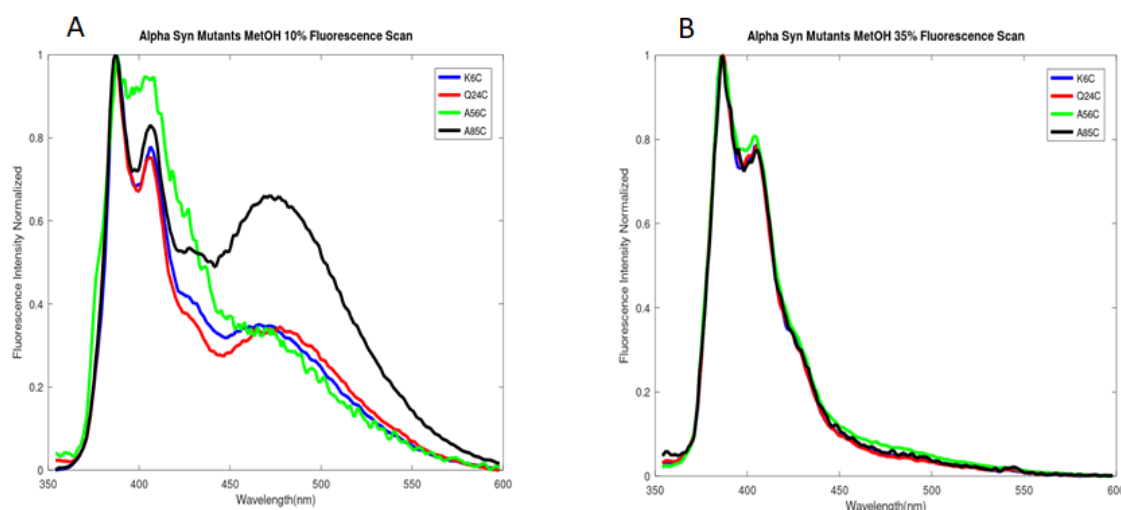

Figure 11. *Comparison between the different pyrene labelled aS mutants. A: Fluorescence emission spectra of pyrene-labelled aS mutants' aggregates at MetOH 10% aggregation regime. B: Fluorescence emission spectra of pyrene-labelled aS mutants aggregates at MetOH 35% aggregation regime. Each plot is representative of a triplicate.*

The observed differences in E/M ratios could be attributed to differences in orientation or rotational freedom of pyrene in the different mutants, which could cause that the pyrene molecules would not exhibit the proper orientation for optimal excimer formation upon parallel aggregation.

On the other hand, the fibrils obtained at high MetOH concentration (MetOH 35%) show a hardly visible excimer band in the emission fluorescence spectra, as can be observed in Figure 11B. This would imply that pyrene rings in neighbor monomers within the aggregates are far apart from each other and this leads to the conclusion that these aggregates possess an antiparallel configuration.

Therefore, the pyrene labelled αS mutants could be used to tell apart parallel from antiparallel aggregates. Based on the emission signal intensities and the E/M ratio, the most suitable αS mutant to move forward with the experiments was the αS-A85C, since it is the labelling position which offers a larger spectral difference between both types of aggregates. The fibrils formed by aggregation of this protein variant exhibited the highest E/M ratio upon parallel aggregation conditions (MetOH 10%) and under antiparallel aggregation regime (MetOH 35%) the excimer band could not be observed (Figure 12).
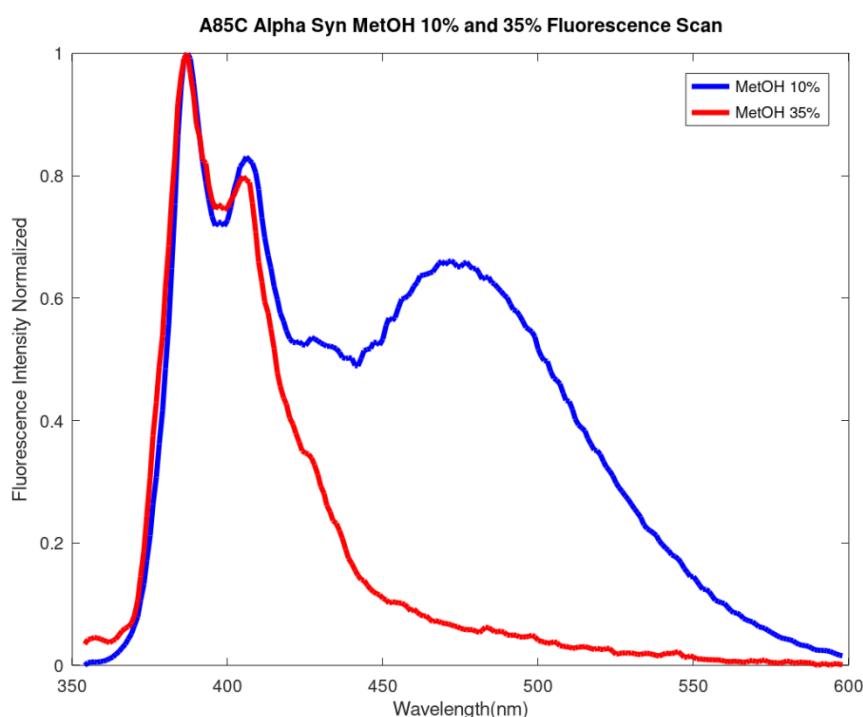

Figure 12. *Fluorescence emission spectra of the fibrils formed by the aggregation of αS-A85C-Pyr under parallel and antiparallel regimes.*

Quantification of the fraction of parallel and antiparallel aggregates under relevant MetOH concentration conditions

Aggregation of the pyrene labelled αS-A85C variant (αS-A85C-Pyr) was induced by using a ratio of c(αS-A85C) : c(αS-A85C-Pyr) = 4:1 since our research group had already noticed that by using this proportion, changes in the aggregation pathway are not induced due to the presence of the probe. MetOH concentrations of 10%, 15%, 16.5%, 18%, 20%, 22.5%, 25%

and 35% were used and triplicates were prepared for each condition. By using this range of MetOH concentrations we expected to retrieve results to compare with the ones obtained by FTIR.

The fluorescence emission spectra resulting from the fibrils at the different MetOH concentrations showed an evident step-wise change in the E/M ratios as can be illustrated in Figure 13.
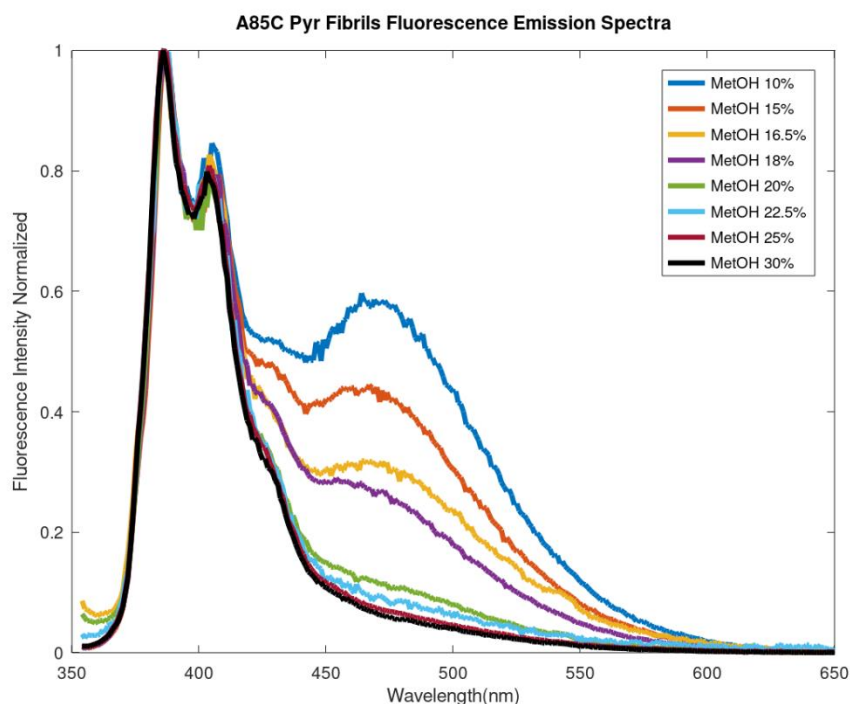

Figure 13. *Fluorescence emission spectra normalized to 386nm of aS-A85C-Pyr aggregates at different MetOH concentrations. Each plot is representative of a triplicate.*

As can be observed in the previous graph, there is a significant increase in the E/M ratio as MetOH concentration decreased, which points to an "in-register" orientation of the ß-sheets in the fibrils core region at lower MetOH concentrations, leading to the subsequent formation of the excimer species by proximal pyrene rings and therefore the appearance of the excimer band at 470nm.

In contrast, at higher MetOH concentrations the pyrene molecules are too far from each other and the probability for two pyrene rings to interact is much lower, which is evidenced by the gradual disappearance of the excimer band in the emission spectra and points to an antiparallel geometry of the fibrils resulting from the aggregation under these conditions.

The averaged E/M ratios and the normalized ratio values were plotted for each aggregation condition studied (Figure 14).

The E/M ratios displayed a nonlinear behavior along the MetOH concentrations studied range, and remained constant for MetOH concentrations higher than 25%. This was

indicative that, at these MetOH concentration values the aggregated fibrils exhibited a 100% antiparallel configuration.
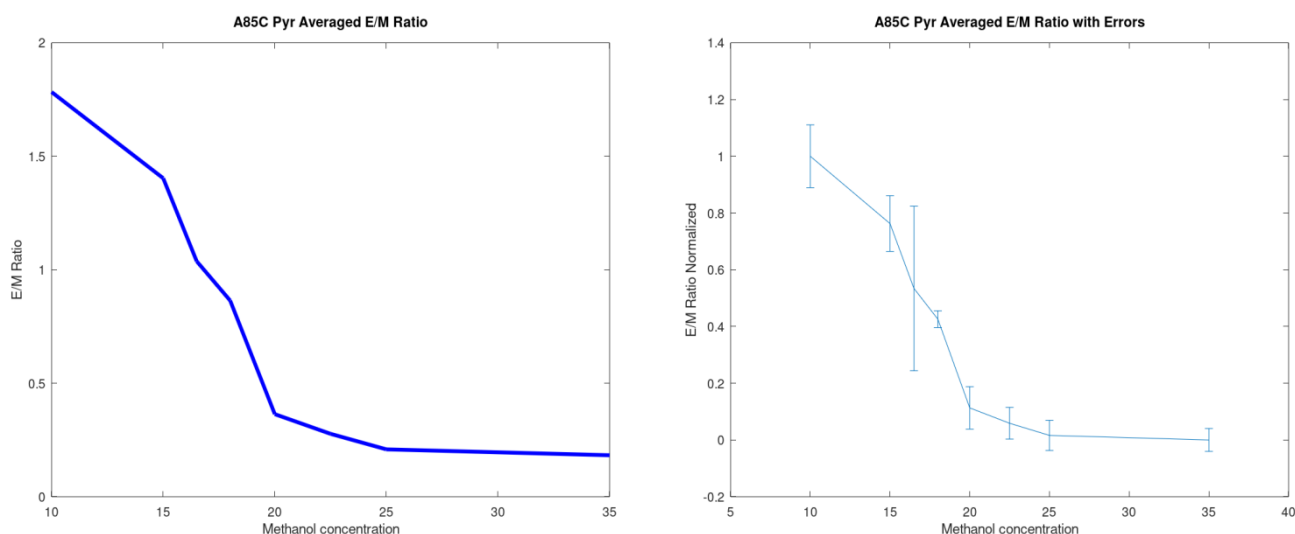

Figure 14. *aS-A85C-Pyr averaged E/M ratio and aS-A85C-Pyr normalized averaged E/M ratio with errors for different MetOH concentration.*

The retrieved values of E/M ratios and the fibrils fraction adopting parallel geometry were represented in Table 2.

Table 2. *Fraction of aS-A85C-Pyr aggregates adopting a parallel geometry.*

MetOH concentration	E/M ratio	E/M Ratio Errors	Parallel αS-A85C-Pyr aggregates fraction
10	1.8	0.1	1.00
15	1.4	0.1	0.76
16.5	1.0	0.3	0.53
18	0.86	0.03	0.43
20	0.36	0.07	0.11
22.5	0.28	0.06	0.06
25	0.21	0.05	0.02
35	0.18	0.04	0.00

In order to provide the αS-A85C-Pyr parallel fractions values for the aggregated fibrils we assumed a 100% of the fibrils adopting an antiparallel conformation at MetOH concentrations equal or higher than 35% and a 100% of the fibrils adopting a parallel geometry at MetOH concentrations equal or lower than 10%. This was also based on the previous results obtained with αS-WT by FTIR-based structural analysis and kinetics experiments, which provided complementary information on the transition between the two aggregation regimes. In particular the inverse β index ratio retrieved by FTIR provides similar information than the E/M for different MetOH conditions as can be observed in Figure 15.
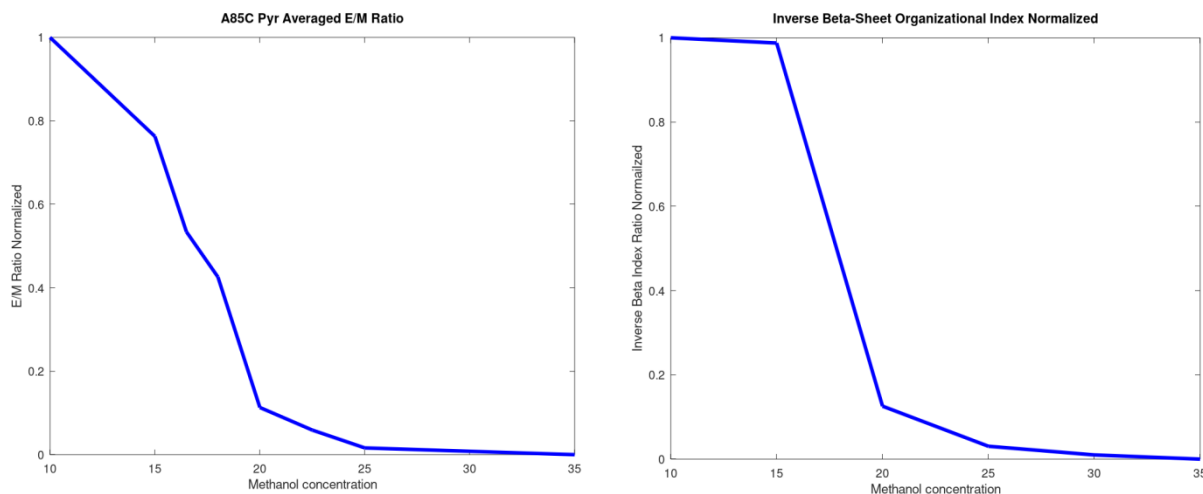
Figure 15. *aS-A85C-Pyr normalized averaged E/M ratio retrieved by fluorescence emission and inverse β index ratio retrieved by FTIR respectively.*

In Figure 15 it can be observed that both approaches, namely pyrene-based fluorescence spectroscopy and FTIR yielded very similar results, which also justified the differences observed in the kinetics experiments for the different MetOH concentrations. It could be observed a small discrepancy between the two observed profiles in Figure 15 for values for MetOH concentrations lower than 15%, since it was expected that at these MetOH concentrations the aggregates would remain 100% parallel according to the inverse β index ratio obtained by FTIR analysis. This could be attributed to fundamental differences in the techniques. For example, the β index ratio is not as rigorous as performing deconvolution of the IR bands in order to obtain a quantification of the secondary structural content. In addition, the 1695 cm^{-1} band in FTIR has typically very low intensity, which could be masked within the noise of the technique at conditions of low MetOH concentrations in the intermediate regime.

From the previous results we could assert that at MetOH concentration values higher than 25% the predominant aggregation pathway of αS is defined by an antiparallel relative monomer orientation and for MetOH concentrations lower than 15% the aggregation of αS takes place mainly through a parallel pathway. At intermediate MetOH concentrations, the two pathways co-exist and a mixture of both types of aggregates is observed. Interestingly, the results show that the transition from a regime of formations of parallel β-sheet aggregates to a regime where anti-parallel β-sheet aggregates are preferred occurs within a very narrow range of MetOH concentrations as in typical phase transition systems.

Determination of a reliable secondary-structure propensity prediction

As explained in Methods, the ROC curves for the predictors NetsurfP2.0, Jpred, Agadir and Sopma were retrieved using 26 proteins with known PDB files.
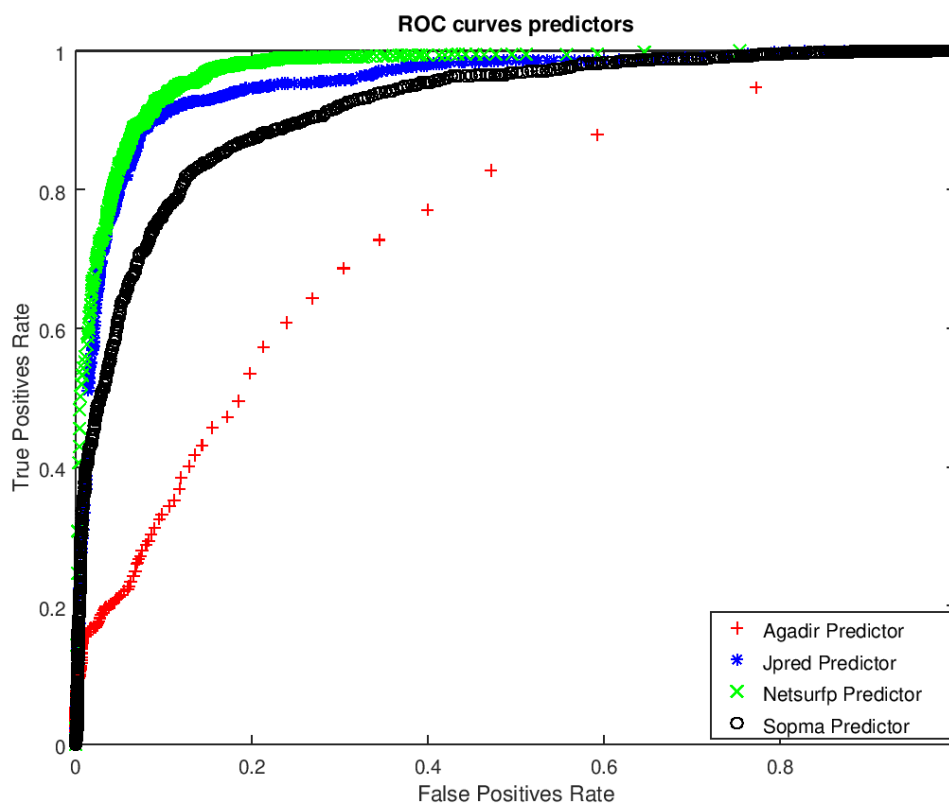

Figure 16. *ROC curves for online predictors.*

As it can be observed in Figure 16, the best predictor for alpha helix conformation was Netsurfp2.0, as it yielded the largest value of area under the curve, of about 0.971. The threshold value that yields the most accurate predictions was 0.6055 ± 0.0003 which corresponds with the largest value of Youden's index of 0.83637 and an accuracy index of 0.91543 (we have found that Youden's index and accuracy index substantially agree in the identification of the threshold value). Residues predicted to be helical with a probability score higher the the threshol value are highly likely to be in a helical conformation in the deposited pdb structure, whereas the prediction values lower or equal to 0.6055 are highly likely not to be helical residues using the Netsurfp2.0 predictor. It should be noticed that this analysis was performed only to determine the reliability of the alpha helix conformation prediction.

Then, the αS sequence was submitted to NetsurfP2.0 predictor, and the results are reported in Figure 17: according to the threshold value identified above there are two well established cores of residues with high alpha helix propensity: one for residues 13 to 30 and the other for residues 56 to 90. As can be observed there are relatively low alpha helix probabilities in the residues located in extremes of the sequence and in between the two alpha helix cores.
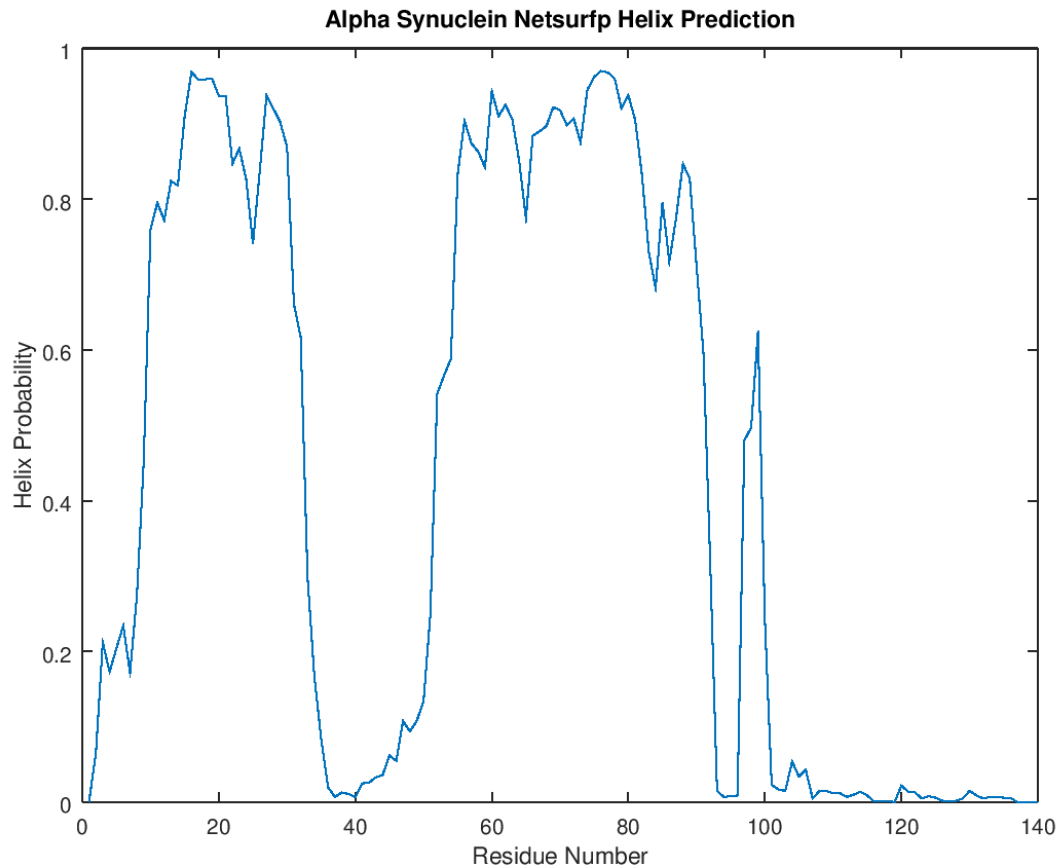
Figure 17. *NetsurfP2.0 alpha helix prediction on αS sequence.*

The region composed by amino acid residues from 56 to 90 coincides with the NAC region of the protein which is known to be highly involved in the β-sheet core in αS parallel aggregates, whereas the amino acid residues from 13 to 30 are located in the N-terminal region of αS sequence.

Based on these results our hypothesis is that amino acids involved in the αS parallel nucleation mechanism are the ones present in the NAC region whereas the amino acids involved in the antiparallel nucleation mechanism are those present in the N-terminal region of the αS sequence.

Conclusions

The work presented in this Master Thesis is part of a bigger project aiming at characterizing the microscopic (at the residue-level) determinants of the complex kinetics and structural polymorphism of alpha-synuclein oligomerization and aggregation. Here, we have mainly approached the experimental characterization of the phase-diagram of the aggregates, as a function of the MetOH concentration, used as a convenient promoter of aggregation, that does not require the presence of interfaces between heterogeneous media.

The presence of structurally different "equilibrium" geometrical arrangements of the aggregates, at different MetOH concentrations, also yields different "kinetic signatures" of the aggregation process, so that different and complementary, experimental techniques have been used, probing different experimental aggregation observables, to provide a robust evidence for the interpretation of the results. In detail:

1. The αS WT kinetics at different MetOH concentrations using ThT fluorescence emission showed an aggregation regime change where a decrease in the lag time was observed when increasing the MetOH concentrations until a point in which a drastic change to a regime with kinetics with no apparent lag-phase was observed.

2. By the structural characterization of the αS amyloid aggregates formed, we could relate the two regimes of αS aggregation kinetics with the formation of aggregates with either parallel or antiparallel β-sheet geometry. The observed transition phase could the established for MetOH concentrations between 15% and 25%, where considerable populations of both amyloid polymorphs coexisted.

3. The pyrene labelling of several αS mutants was carried out and by analyzing the excimer-to-monomer (E/M) ratio of the different pyrene-αS variants we can conclude that Pyr-A85C-αS is the most suitable variant for assessing αS aggregation and for quantifying the fraction of parallel β-sheet aggregates in complex mixtures of different types of aggregates.

4. By pyrene excimer fluorescence emission it could be monitored the αS aggregation and the information retrieved by FTIR and ThT kinetics experiments was complemented arriving to comparable results in the transition phase between parallel and antiparallel regimes.

On the theoretical side, we have moved the first steps towards the modification and application of the WSME-agg model to alpha-synuclein, providing a proposal for the native geometry, based on secondary structure predictions from public servers, rather than using the experimental data from available pdb structures where the protein is strongly stabilized by interactions with other molecules, that could alter the intrinsic propensities. To this end, we have proceeded as follows:

5. The assessment of several online secondary structure predictors was performed and by statistical analysis using ROC curves it was found that the Netsurfp2.0 provided the most accurate results for alpha helix predictions on amino acid sequences.

6. By using the alpha helix predictor Netsurfp2.0 it could be found that in the αS sequence, there are two main regions with high alpha helix propensity, which we believe are associated with the two types of aggregation mechanisms and amyloid polymorphs observed in the presence of MetOH.

Follow up experiments

Several aspects should be further addressed in future work, according at least to the following recommendations:

1. To perform a more detailed kinetics analysis on αS aggregation reaction in the transition phase between parallel and antiparallel regimes.
2. To optimize the reaction conditions of αS pyrene maleimide in order to enhance the reaction yields.
3. To label other positions in the αS sequence this time based on the predictions of the NeSurfp2.0 online predictor and the WSME approach.
4. To use the pyrene labelled αS mutants to study the kinetics of aggregation in the presence of MetOH and at more physiologically relevant conditions.
5. To study the aggregation of pyrene labelled αS mutants by using the pyrene band I/band III ratio.
6. To study the α-helix to β-sheet transition in the αS sequence using Zimm-Bragg model.
7. To simulate the αS aggregation by using the WSME-agg model.

References

1- Chiti, F., Dobson, C.M., Protein misfolding, functional amyloid, and human disease, 2006, Annu. Rev. Biochem. 75, 333–366.

2- Eisenberg, D., Jucker, M., The amyloid state of proteins in human diseases, 2012, Cell 148 1188–1203.

3- Fandrich, M., Dobson, C.M., The behaviour of polyamino acids reveals an inverse sidechain effect in amyloid structure formation, 2002, EMBO J. 21, 5682–5690.

4- Nunilo Cremades, Christopher M. Dobson, The contribution of biophysical and structural studies of protein self-assembly to the design of therapeutic strategies for amyloid diseases, 2018, Neurobiology of Disease 109 178–190.

5- Cremades N, *et al,* Direct observation of the interconversion of normal and toxic forms of α-synuclein, 2012, *Cell* 149(5):1048-1059.

6- Marija Iljinaa, Gonzalo A. Garcia, Mathew H. Horrocksa, Laura Tosatto, Minee L. Choic, Kristina A. Ganzinger, Andrey Y. Abramovc, Sonia Gandhb, Nicholas W. Woodc, Nunilo Cremades, Christopher M. Dobsona, Tuomas P. J. Knowlesa, and David Klenermana, Kinetic model of the aggregation of alpha-synuclein provides insights into prion-like spreading, 2016, PNAS, E1206-E1215.

7- Clayton, D.F., George, J.M., The synucleins: a family of proteins involved insynaptic function, plasticity, neurodegeneration and disease 1998, Trends Neurosci. 21, 249–254.

8- Apetri, M. M., Maiti, N. C., Zagorski, M. G., Carey, P. R. and Anderson, V. E., Secondary structure of α-synuclein oligomers: characterization by Raman and atomic force microscopy., 2006, J. Mol. Biol. 355, 63–71.

9- Giuliana Fusco, Serene W. Chen, Philip T. F. Williamson, Roberta Cascella, Michele Perni, James A. Jarvis, Cristina Cecchi, Michele Vendruscolo, Fabrizio Chiti, Nunilo Cremades, Liming Ying, Christopher M. Dobson, and Alfonso De Simone, Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers, 2017, Science, 358(6369):1440-1443.

10- Chen, M., Margittai, M., Chen, J., Langen, R., Investigation of alpha-synuclein fibril structure by site-directed spin labeling. J. Biol. Chem., 2007, 282, 24970–24979.

11- Vilar, M., Chou, H.T., Luhrs, T., Maji, S.K., Riek-Loher, D., Verel, R., Manning, G.,Stahlberg, H., Riek, R., The fold of alpha-synuclein fibrils, 2008, Proc. Natl. Acad.Sci. U.S.A. 105, 8637–8642.

12- María Soledad CELEJ, Rabia SARROUKH, Erik GOORMAGHTIGH, Gerardo D. FIDELIO* Jean-Marie RUYSSCHAERT and Vincent RAUSSENS, Toxic prefibrillar α-synuclein amyloid oligomers adopt a distinctive antiparallel β-sheet structure, 2012, Biochem. J., 443, 719–726.

13- Ellen Hubin, Stéphanie Deroo, Gabriele Kaminksi Schierle, Clemens Kaminski, Louise Serpell, Vinod Subramaniam, Nico van Nuland, Kerensa Broersen, Vincent Raussens, Rabia Sarroukh., Two distinct b-sheet structures in Italian-mutant amyloid-beta fibrils: a potential link to different clinical phenotypes., 2015, Cell. Mol. Life Sci., 72, 4899–4913.

14- Steven J. Roeters, Aditya Iyer, Galja Pletikapić, Vladimir Kogan, Vinod Subramaniam, Sander Woutersen., Evidence for Intramolecular Antiparallel Beta-Sheet Structure in Alpha-Synuclein Fibrils from a Combination of Two-Dimensional Infrared Spectroscopy and Atomic Force Microscopy, 2017, Scientific Reports 7, 41051.

15- Serene W. Chen, Srdja Drakulic, Emma Deas, Myriam Ouberai, Francesco A. Aprile, Rocío Arranz, Samuel Ness, Cintia Roodveldt, Tim Guilliams, Erwin J. De-Genst, David Klenerman, Nicholas W. Wood, Tuomas P.J. Knowles, Carlos Alfonso, Germán Rivas, Andrey Y. Abramov, José María Valpuesta, Christopher M. Dobson, and Nunilo Cremades, Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation, 2015, PNAS 112 (16) E1994-E2003.

16- Sangita Seshadri, Ritu Khurana, Anthony L.Fink, Fourier transform infrared spectroscopy in analysis of protein deposits, 1999, Methods in Enzymology, 309, 559-576.

17- Rabia Sarroukh, Erik Goormaghtigh, Jean-Marie Ruysschaert , Vincent Raussens, ATR-FTIR: A "rejuvenated" tool to investigate amyloid proteins, 2013, Biochimica et Biophysica Acta 1828: 2328-2338.

18- Andreas Barth and Christian Zscherp, What vibrations tell us about proteins, 2002, Quarterly Reviews of Biophysics, 35 369–430.

19- Shyamala Thirunavukkuarasu, Elizabeth A. Jares-Erijman and Thomas M. Jovin, Multiparametric Fluorescence Detection of Early Stages in the Amyloid Protein Aggregation of Pyrene-labeled α-Synuclein, 2008, J. Mol. Biol. 378, 1064-1073.

20- Gursharan K. Bains, Sea H. Kim, Eric J. Sorin, and Vasanthy Narayanaswami, Extent of Pyrene Excimer Fluorescence Emission is a Reflector of Distance and Flexibility: Analysis of the Segment Linking the LDL Receptor-binding and Tetramerization Domains of Apolipoprotein E3[1], 2012, Biochemistry 51(31): 6207–6219.

21- Samuel I. A. Cohen, Michele Vendruscolo, Mark E. Welland, Christopher M. Dobson, Eugene M. Terentjev, and Tuomas P. J. Knowles, Nucleated polymerization with secondary pathways I. Time evolution of the principal moments, 2011, J Chem. Phys., 135(6), 065105.

22- Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles, Nucleated polymerization with secondary pathways II. Determination of self-consistent solutions to growth processes described by non-linear master equations, 2011, J Chem. Phys., 135(6), 065106.

23- Wako, H.; Saito, N., Statistical Mechanical Theory of the Protein Conformation. I. General Considerations and the Application to Homopolymers, 1978,J. Phys. Soc. Jpn., 44, 1931–1938.

24- Wako, H.; Saito, N., Statistical Mechanical Theory of the Protein Conformation. II. Folding Pathway for Protein, 1978, J. Phys. Soc. Jpn., 44, 1939–1945.

25- Muñoz, V. M.; Eaton, W. A. Proc. Natl. Acad. Sci. U. S. A. 1999, 96, 11311–6.

26- Bruscolini, P.; Pelizzola, A. Phys. Rev. Lett. 2002, 88, 258101.

27- M. Zamparo, A. Trovato, and A. Maritan, Simplified Exactly Solvable Model forß-Amyloid Aggregation, 2010, PHYSICAL REVIEW LETTERS 105.108102.

28- Christine Xue, Tiffany Yuwen Lin, Dennis Chang and Zhefeng Guo, Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation, 2017, R. Soc. 4: 160696.

29- Matthew Biancalana and Shohei Koide, Molecular Mechanism of Thioflavin-T Binding to Amyloid Fibrils, 2010, Biochim Biophys Acta, 1804(7): 1405–1412.

30- Howard Siu and Jean Duhamel, Molar Absorption Coefficient of Pyrene Aggregates in Water, 2008, J. Phys. Chem., 112, 15301–15312.
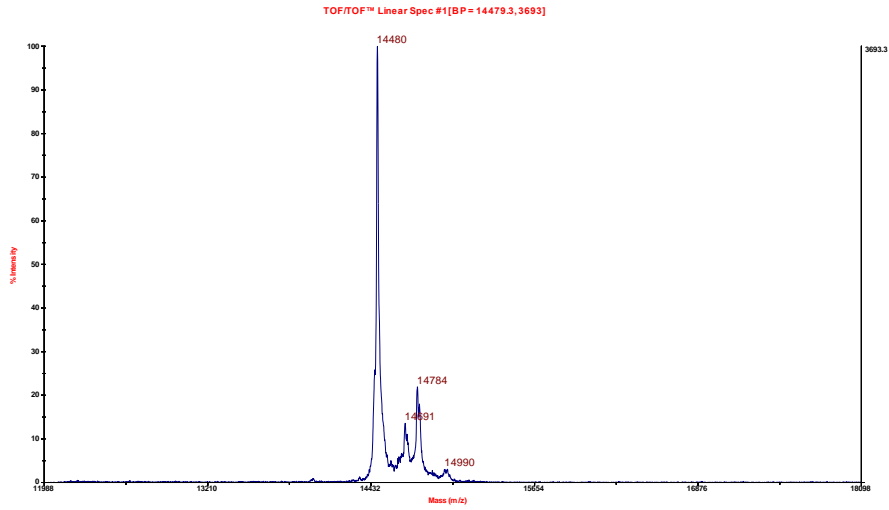
Appendixes

Appendix 1: *Maldi MS Spectra of pyrene labelled aS mutants*
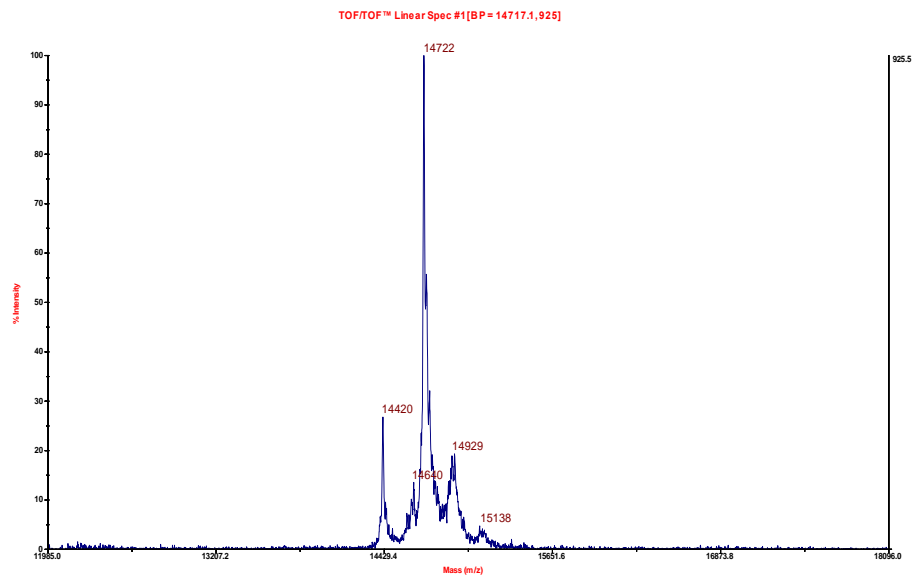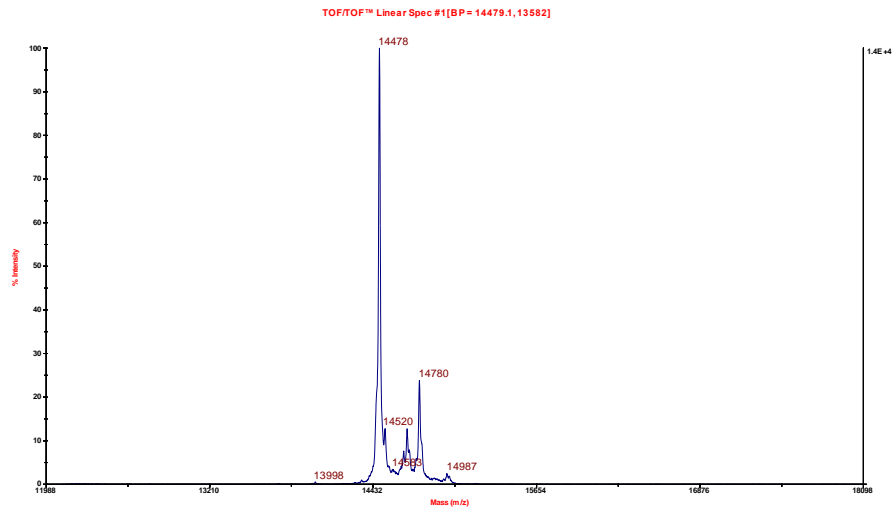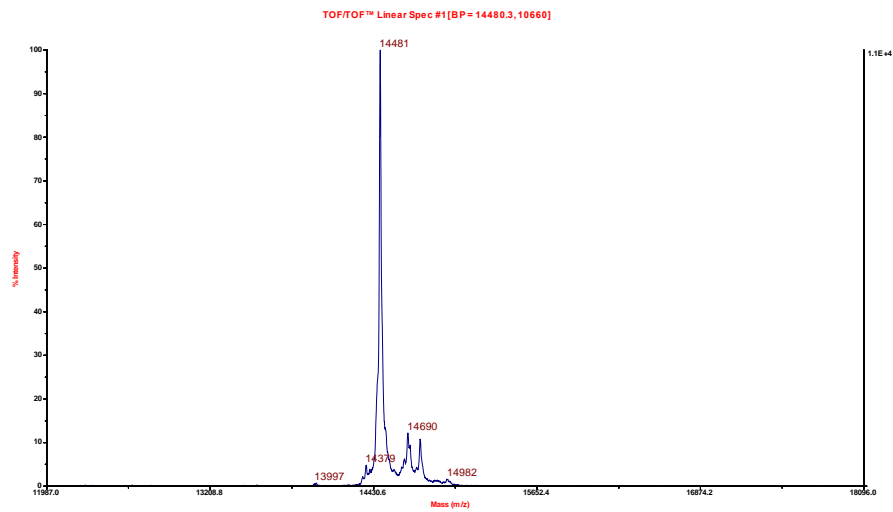

Appendix 1A: *aS Pyr-K6C Mass spectrum*
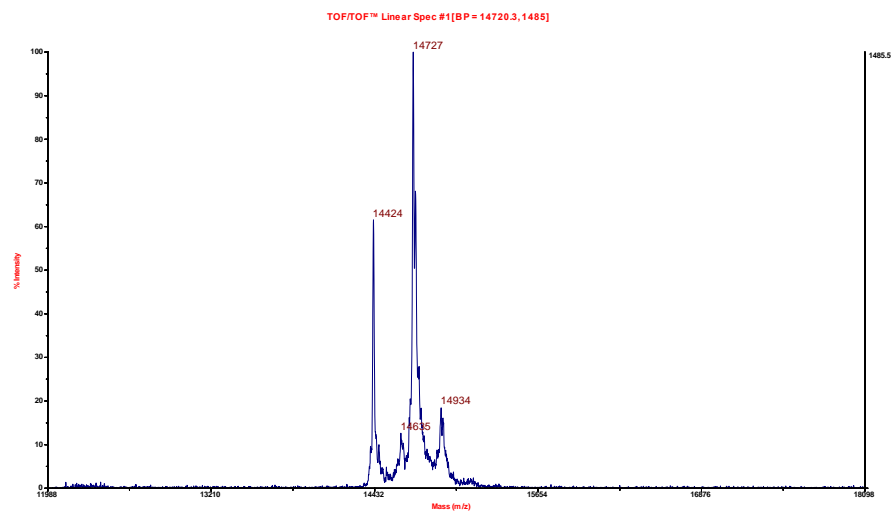
Appendix 1B: *aS Pyr-Q24C Mass spectrum*

Appendix 1C: *aS Pyr-A56C Mass spectrum*

Appendix 1C: *aS Pyr-A85C Mass spectrum*

Appendix 1D: *aS Pyr-A140C Mass spectrum*

Appendix 2: *Script for online secondary structure predictors ROC curve*

```
1;
clear all;
close all;

function [TPR, FPR, area, Youden, bestthreshold, accindexv] = ROC(threshold, dssout, predout)
TPR=zeros((length(threshold)),1);
FPR=zeros((length(threshold)),1);
Youden=zeros((length(threshold)),1);
accindex=zeros((length(threshold)),1);
for i=1:length(threshold)
  TP=zeros(length(threshold),1);
  TN=zeros(length(threshold),1);
  FP=zeros(length(threshold),1);
  FN=zeros(length(threshold),1);
  for j= 1:length(predout)
    if ((dssout(j) == "H") && (predout(j) > threshold(i)))
    TP(i) = TP(i) + 1;
  elseif ((dssout(j) =="H") && (predout(j) <= threshold(i)))
    FN(i) = FN(i) + 1;
  elseif ((dssout(j) != "H") && (predout(j) <= threshold(i)))
    TN(i) = TN(i) +1;
  else
    FP(i) = FP(i) +1;
    endif
  endfor
  FPR(i) = FP(i)/(FP(i) + TN(i));
  TPR(i) = TP(i)/(TP(i) + FN(i));
  accindex(i)= (TP(i)+TN(i))/(TP(i)+TN(i)+FP(i)+FN(i));
endfor

## Area computation

ar=[];
for k=1:(length(threshold)-1)
  #ar(k)=(abs(FPR(k)-FPR(k+1))*TPR(k));
  ar(k)=(FPR(k)-FPR(k+1))*(TPR(k)+TPR(k+1))/2.;
endfor
#ar
area=sum(ar);

#Distance computation
for i=1:length(threshold)
  Youden(i)=(TPR(i)-FPR(i));
endfor
```

```
maxyou=max(Youden);
index=find(Youden==maxyou);
bestthreshold=threshold(index);
accindexv=accindex(index);
endfunction
```

#Load the variables

```
[a, agout]= textread("total_agadir.txt","%s %f");
[c, soout]= textread("total_sopma.txt","%s %f");
[e, neout]= textread("total_netsurfp.txt","%s %f");
[g, jpout]= textread("total_jpred.txt","%s %f");
[l, m]= textread("total_dssp.txt","%s %s");
```
#Declare as ags,jps,nes, sos and dss the sequences reported in the prediction files and the
sequence reported by dssp respectively
```
ags=[]
for i=1:length(a)
  ags=[ags;a{i}(1)];
endfor

sos=[]
for i=1:length(c)
 sos=[sos;c{i}(1)];
endfor

nes=[]
for i=1:length(e)
  nes=[nes;e{i}(1)];
endfor

jps=[]
for i=1:length(g)
  jps=[jps;g{i}(1)];
endfor

dss=[]
for i=1:length(g)
  dss=[dss;l{i}(1)];
endfor
```

#Declare as dssout the outcome of the reported secondary structure for the sequences
```
dssout=[]
 for i=1:length(m)
 dssout=[dssout;m{i}(1)];
endfor
```

# Declare a complementary vector for the determination of the minimal non zero value of each predictor output and retrieve the minimal and maximal values
agaux=agout;
#agaux(agaux==0)= [];
minag= min(agaux);
maxag= max(agaux);

neaux=neout;
#neaux(neaux==0)=[];
minne=min(neaux);
maxne=max(neaux);

soaux=soout;
#soaux(soaux==0)=[];
minso=min(soaux);
maxso=max(soaux);

jpaux=jpout;
#jpaux(jpaux==0)=[];
minjp=min(jpaux);
maxjp=max(jpaux);

#Thresholds for each predictor: 5% lower than minimal value and 5% higher than maximal value
#thresag, thresso, thresne and thresjp for agadir, sopma , netsurfp and jpred

thresag=linspace((minag-0.05*maxag),(maxag+0.05*maxag), 5000);
thresso=linspace((minso-0.05*maxso),(maxso+0.05*maxso), 5000);
thresne=linspace((minne-0.05*maxne),(maxne+0.05*maxne), 5000);
thresjp=linspace((minjp-0.05*maxjp),(maxjp+0.05*maxjp), 5000);

 # Check length missmatching among predictors and dssp files
 if ((length(a) != length(c)) || (length(a) != length(c)) || (length(a) != length(e)) ||(length(a) != length(g)))
   disp("different lengths sequences between predictors!");
 endif
 if (length(a)!=length(l))
   disp("different lengths sequences between predictor agadir and dssp file!");
 else
   disp("agadir predictor file has the same length as the dssp file");
 endif
 if (length(c)!=length(l))
   disp("different lengths sequences between predictor sopma and dssp file!");
 else
   disp("sopma predictor file has the same length as the dssp file");
 endif

```
  if (length(e)!=length(l))
    disp("different lengths sequences between predictor netsurfp and dssp file!");
  else
    disp("netsurfp predictor file has the same length as the dssp file");
  endif
  if (length(g)!=length(l))
    disp("different lengths sequences between predictor jpred and dssp file!");
  else
    disp("jpred predictor file has the same length as the dssp file");
  endif

  #Check residue missmatching between predictors and dssp files
  for i=1:length(dss)
if (strcmp(dss(i),ags(i)) != 1)
  index=i;
  #dss(i)
  #ags(i)
  disp("for agadir predictor missmatching sequences in residue:");
  disp(index);
endif
endfor

for i=1:length(dss)
if (strcmp(dss(i),sos(i)) != 1)
  index=i;
  #dss(i)
  #sos(i)
  disp(" for sopma predictor missmatching sequences in residue:");
  disp(index);
endif
endfor

for i=1:length(dss)
if (strcmp(dss(i),nes(i)) != 1)
  index=i;
  #dss(i)
  #nes(i)
  disp(" for netsurfp predictor missmatching sequences in residue:");
  disp(index);
endif
endfor
for i=1:length(dss)
if (strcmp(dss(i),jps(i)) != 1)
  index=i;
  #dss(i)
  #jps(i)
  disp(" for jpred predictor missmatching sequences in residue:");
```

```
  disp(index);
endif
endfor


#Variables initialization for computing the ROC values for Agadir predictor.
TPRag=zeros((length(thresag)),1);
FPRag=zeros((length(thresag)),1);
youdenag=zeros((length(thresag)),1);
areaag=0;
accindexag=0;
bestthresholdag=0;
#Computing ROC values for Agadir predictor
[TPRag FPRag areaag  youdenag bestthresholdag accindexag] = ROC(thresag,dssout,agout);
disp("The best threshold for Agadir predictor is:")
disp(bestthresholdag);
disp("The area under the curve for Agadir predictor is:")
disp(areaag);
disp("The accuracy index for Agadir predictor is:")
disp(accindexag);



#Variables initialization for computing the ROC values for Sopma predictor.
TPRso=zeros((length(thresso)),1);
FPRso=zeros((length(thresso)),1);
youdenso=zeros((length(thresso)),1);
areaso=0;
bestthresholdso=0;
accindexso=0;
#Computing ROC values for Sopma predictor
[TPRso FPRso areaso youdenso bestthresholdso accindexso] = ROC(thresso,dssout,soout);
disp("The best threshold for Sopma predictor is:")
disp(bestthresholdso);
disp("The area under the curve for Sopma predictor is:")
disp(areaso);
disp("The accuracy index for Sopma predictor is:")
disp(accindexso);



#Variables initialization for computing the ROC values for Netsurfp predictor.
TPRne=zeros((length(thresne)),1);
FPRne=zeros((length(thresne)),1);
youdenne=zeros((length(thresne)),1);
areane=0;
bestthresholdne=0;
accindexne=0;
#Computing ROC values for Sopma predictor
[TPRne FPRne areane youdenne bestthresholdne accindexne] = ROC(thresne,dssout,neout);
```

```
disp("The best threshold for Netsurfp predictor is:")
disp(bestthresholdne);
disp("The area under the curve for Netsurfp predictor is:")
disp(areane);
disp("The accuracy index for Netsurfp predictor is:")
disp(accindexne);


#Variables initialization for computing the ROC values for Jpred predictor.
TPRjp=zeros((length(thresjp)),1);
FPRjp=zeros((length(thresjp)),1);
youdenjp=zeros((length(thresjp)),1);
areajp=0;
bestthresholdjp=0;
accindexjp=0;
#Computing ROC values for Sopma predictor
[TPRjp FPRjp areajp youdenjp bestthresholdjp accindexjp] = ROC(thresjp,dssout,jpout);
disp("The best threshold for Jpred predictor is:")
disp(bestthresholdjp);
disp("The area under the curve for Jpred predictor is:")
disp(areajp);

disp("The accuracy index for Jpred predictor is:")
disp(accindexjp);

comparacion=figure(1)

plot(FPRag,TPRag,'r+');
title("ROC curves predictors");
xlabel("False Positives Rate");
ylabel("True Positives Rate");
hold on;
plot(FPRjp,TPRjp,'b*');
hold on;
plot(FPRne,TPRne,'gx');
hold on;
plot(FPRso,TPRso,'ko')
hold off
legend("Agadir     Predictor","Jpred     Predictor","Netsurfp     Predictor","Sopma
Predictor",'Location','SouthEast');
#saveas(comparacion,"predictores2.tiff");
```

Appendix 3: *Linux bash scripts used in the manipulation of the data retrieved from the online predictors.*

A file named 'list' was created with the names of the pdb files of the studied proteins in a column (for example 1a5e)

The pdb files were downloaded by executing the following bash script with the 'list' file in the same directory:

```
#!/bin/bash

for a in $(cat list)

do

  wget https://files.rcsb.org/download/$a.pdb

done
```

Once we had the pdb files, the dssp files were retrieved from them by using the bash script:

```
#!/bin/bash

for a in $(cat list)

do

  mkdssp $a.pdb -o $a.dssp

done
```

The sequence submissions to the online predictors had to be done one by one.

DSSP values:

The values of residue code and alpha helix prediction were copied to a single file:

```
awk '{print($4,$5)}' 1a5e.dssp 1a23.dssp 1ayf.dssp 1bni.dssp 1bpi.dssp 1bta.dssp 1c9o.dssp 1chk.dssp 1cun.dssp 1cyo.dssp 1fkj.dssp 1ftg.dssp 1gd1.dssp 1h7m.dssp 1hfy.dssp 1otr.dssp 1pga.dssp 1rhg.dssp 1rx4.dssp 1sce.dssp 1stn.dssp 2ci2.dssp 2rn2.dssp 2trx.dssp 3ssi.dssp 451c.dssp > total_dssp.txt
```

Agadir predictor:

The column 3 retrieved the alpha helix percentage so it was converted to probability and all preditions were copied to a single file:

```
 awk '{print($2,$3/100)}' 1a5e_agadir.txt 1a23_agadir.txt 1ayf_agadir.txt 1bni_agadir.txt 1bpi_agadir.txt   1bta_agadir.txt   1c9o_agadir.txt   1chk_agadir.txt   1cun_agadir.txt 1cyo_agadir.txt   1fkj_agadir.txt   1ftg_agadir.txt   1gd1_agadir.txt   1h7m_agadir.txt
```

1hfy_agadir.txt   1otr_agadir.txt   1pga_agadir.txt   1rhg_agadir.txt   1rx4_agadir.txt
1sce_agadir.txt   1stn_agadir.txt   2ci2_agadir.txt   2rn2_agadir.txt   2trx_agadir.txt
3ssi_agadir.txt 451c_agadir.txt > total_agadircomas.txt

The commas in the Agadir file were removed:

awk '{ gsub(/,/,""); print }' total_agadircomas.txt > total_agadir.txt

Jpred predictor:

Part of the Jpred data treatment was performed manually.

The file containing all the Jpred predictions was created:

awk 'NF>0' 1a5e_jpred.txt 1a23_jpred.txt 1ayf_jpred.txt 1bni_jpred.txt 1bpi_jpred.txt
1bta_jpred.txt 1c9o_jpred.txt 1chk_jpred.txt 1cun_jpred.txt 1cyo_jpred.txt 1fkj_jpred.txt
1ftg_jpred.txt   1gd1_jpred.txt   1h7m_jpred.txt   1hfy_jpred.txt   1otr_jpred.txt
1pga_jpred.txt 1rhg_jpred.txt 1rx4_jpred.txt 1sce_jpred.txt 1stn_jpred.txt 2ci2_jpred.txt
2rn2_jpred.txt 2trx_jpred.txt 3ssi_jpred.txt 451c_jpred.txt > total_jpred.txt

NetsurfP2.0 predictor:

A previous manual treatment was performed on the files where the unwanted comments were removed. The file columns 2 and 8 were copied where the column 2 contained the aminoacid code and the column 8 contained the alpha helix propensities, this was performed for all prediction files:

awk '{print ($2, $8)}'

The files were afterwards copied to a single directory:

#!/bin/bash

for a in $(cat lista.txt)

do

  cd $a/Archivos_procesados

  cp $a"_netsurfp".txt ../../Total_netsurfp

  cd ../..

done

The files were combined in one in the same way as above.

Sopma predictor:

The retrieved values from this predictor had to be normalized:

```
awk '{print($2)/(($2)+($3)+($4)+($5))}' SOPMA\ columns.txt > SOPMA_ALphaSynHelixProb2.txt
```

The files were then copied to a single directory:

```
#!/bin/bash

for a in $(cat lista.txt)

do

  cd $a/Archivos_procesados

  cp $a"_sopma".txt ../../Total_Sopma

  cd ../..

done
```

The file containing all predictions was created by using the bash script:

```
#!/bin/bash

for a in $cat(lista.txt)

do

awk '{NF>0}' total_sopma $a"_sopma".txt > total_sopma

done
```