UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Doctoral Thesis

Innovative Methods to Investigate Intrinsically Disordered Proteins by Small-Angle Scattering Techniques

PhD Student:
Paolo Moretti

Supervisor:
Prof. Francesco Spinozzi

Co-Tutor:
Prof. Paolo Mariani

Department of Life and Environmental Sciences - Di.S.V.A.

XXXI cycle
Ancona, 2017-2018

Declaration of Authorship

I, Paolo Moretti, declare that this thesis titled, *Innovative Methods to Investigate Intrinsically Disordered Proteins by Small-Angle Scattering Techniques* and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

UNIVERSITÀ POLITECNICA DELLE MARCHE

Abstract

Department of Life and Environmental Sciences - Di.S.V.A.

Doctor of Philosophy

Innovative Methods to Investigate Intrinsically Disordered Proteins by Small - Angle Scattering Techniques

by Paolo MORETTI

My PhD project is focused on the study of the characterization of the conformational properties of Intrinsically Disordered Proteins (IDPs) investigated by Small Angle X-rays or Neutrons Scattering. IDPs are a class of proteins that, despite the absence of a defined tertiary structure, are able to perform a biological function and play an important role in a large number of pathologies. In this thesis I have studied one of the most important proteins involved in neurodegenerative disease worldwide, α - synuclein (α-syn), which is related to Parkinson's diseases. This protein is prone to originate amyloids fibrils as a result of β aggregation processes, which start from the monomeric state and end with the production of amyloid fibrils in Lewy's bodies, the main hallmarks of this dementia [1].

The goal of my PhD project is to describe the early stages of the aggregation kinetics of α-syn in order to understand the mechanisms at the origin of the pathological process.

The novel method discussed in this thesis is based on the combination of computational techniques and experimental measurements that I performed at ESRF, the European Synchrotron (Grenoble), on diluted water solutions of IDPs. Disordered structures, originating from molecular dynamics simulation or related techniques [2], are taken into account by estimating the relative population weights from sets of experimental SAXS data in the framework to the Bayesian formalism. As a result, the developed method allows to quantify the conformational disorder of IDPs in different chemical-physical conditions.

Contents

List of Figures

List of Tables

Abbreviations

IDPs	***I**ntrinsically **D**isordered **P**roteins*
LSF	***L**arge **S**cale **F**acilities*
SAS	***S**mall **A**ngle **S**cattering*
SAXS	***S**mall **A**ngle **X**-Ray **S**cattering*
SANS	***S**mall **A**ngle **N**eutron **S**cattering*
PD	***P**arkinson 's **D**isease*
WT	***W**ild **T**ype*
Rg	***R**adius of **G**yration*
NPs	***N**ano**P**article**s***

Symbols

Å	Angström
ρ	Scattering Length Density
$P(Q)$	Form factor
$S(Q)$	Structure factor

Chapter 1

Introduction

This thesis presents a model of study for the characterization of the conformational properties of Intrinsically Disordered Proteins (IDPs) based on the combinations of computational technique and experimental measurements of Small Angle X-Rays or Neutrons Scattering (SAXS and SANS, respectively) performed at Large Scale Facilities (LSF). The protein studied is α - synuclein (α-syn), which is implicated in Parkinson's Disease (PD). This kind of diseases are classified as neurodegenerative diseases and have as a common feature and as main hallmark the formation of amyloid fibrils in the brain tissue. The formation of amyloid fibrils is caused by a process of aggregation of IDPs that follows an aggregation kinetics starting from the protein in the monomeric state, until the complete formation of the fibrils. The goal of the project is to try to develop a method that makes it possible to investigate these processes with Small Angle Scattering (SAS), from the early stages, trying to understand which conformational changes act as trigger for protein aggregation. The developed method can be applied to different types of IDPs beyond

those shown in this work. The main characteristics of IDPs, neurodegenerative diseases and related proteins will be discussed in this thesis. Finally, the novel method and its applications will be described. The experimental sessions through which the SAXS data of this thesis have been collected were carried out at ESRF, the European Synchrotron, in Grenoble, France.

1.1 Intrinsically Disordered Proteins

The central **structure - function** paradigm of molecular biology states that a protein before being biologically active must fold into a three-dimensional structure, which makes it able to perform precise functions [19]. The information that allows a protein to take a certain fold under physiologic conditions is written in its amino acid sequence [20]. The folding process that can be spontaneous or can be assisted by particular molecules called molecular chaperones, which help the protein to find a precise folding. Despite this sort of dogma is still true, in the last decades the presence of protein that are able to have a biological function without a well-defined structure has become increasingly evident. This kind of proteins have been called Intrinsically Disordered Proteins (Fig. 1.1).

FIGURE 1.1: Order and Disorder [3]

IDPs have been found in many genomes, including human genome, and play important roles in central cellular processes, such as cell cycle and cell signaling and are often involved in diseases, including cancers and neurodegenerative disorders. Since the first studies, it was immediately evident that the IDPs were highly present in every organism and, in particular, in complex organism (Fig. 1.2). Indeed, in bacterial genomes only 4% of all protein are predicted to have disordered regions, whereas in eukaryotes this occurs for around one-third of protein molecules and for humans for around one-half

of them. IDPs are involved in important functions such as signaling hubs in protein - protein interaction, in transport lipids and cholesterol in blood plasma and in many other cases [21].

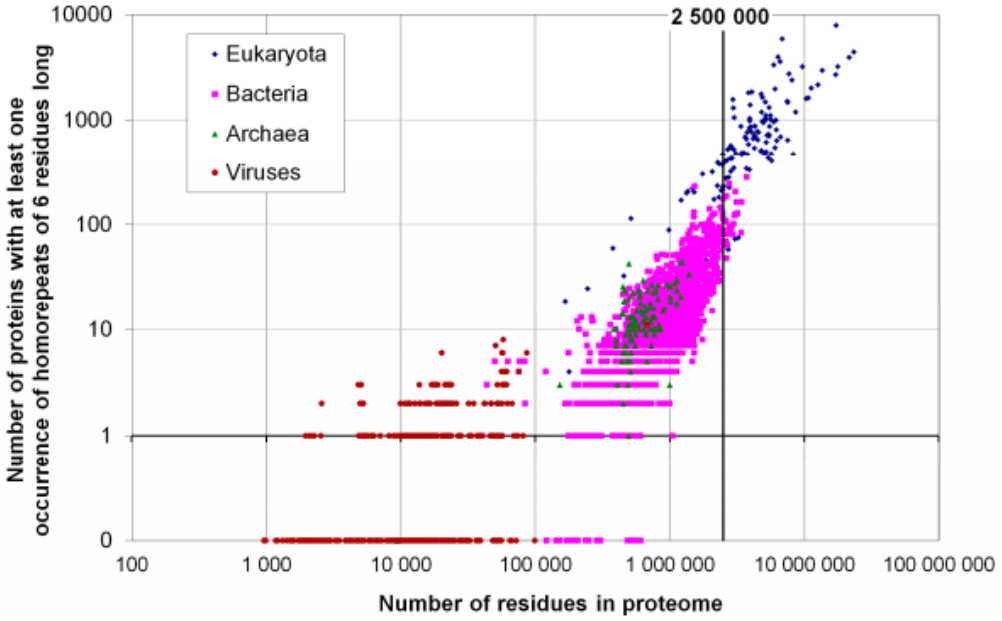


FIGURE 1.2: IDPs in proteome [4]

These facts have attracted much attention from researchers and there has been a large increase in research efforts in this area, as shown in Fig. 1.3.

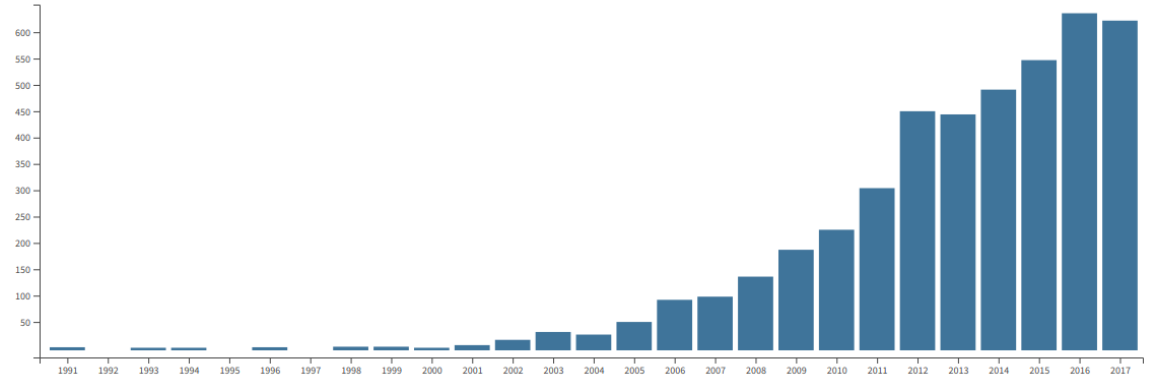


FIGURE 1.3: IDPs publication per years from Web of Science

1.1.1 Looking from Inside

The first step in studying IDPs is the evaluation of their the primary sequence with the aim of understanding if, already at this level, there are differences with the ordered proteins. Analyzing the primary sequence of IDPs, two important difference have been found in the average properties of the amino acids composing them:

- low hydrophobicity;
- high net charge.

Hydrophobicity is one of the driving forces for the compaction of a protein, and the high net charge leads to great electrostatic repulsion forces: the presence of these conditions is certainly an important requirement for the absence of an ordered structure. These differences are well shown in Fig. 1.4, in which the two types of proteins form two different populations based on these properties.

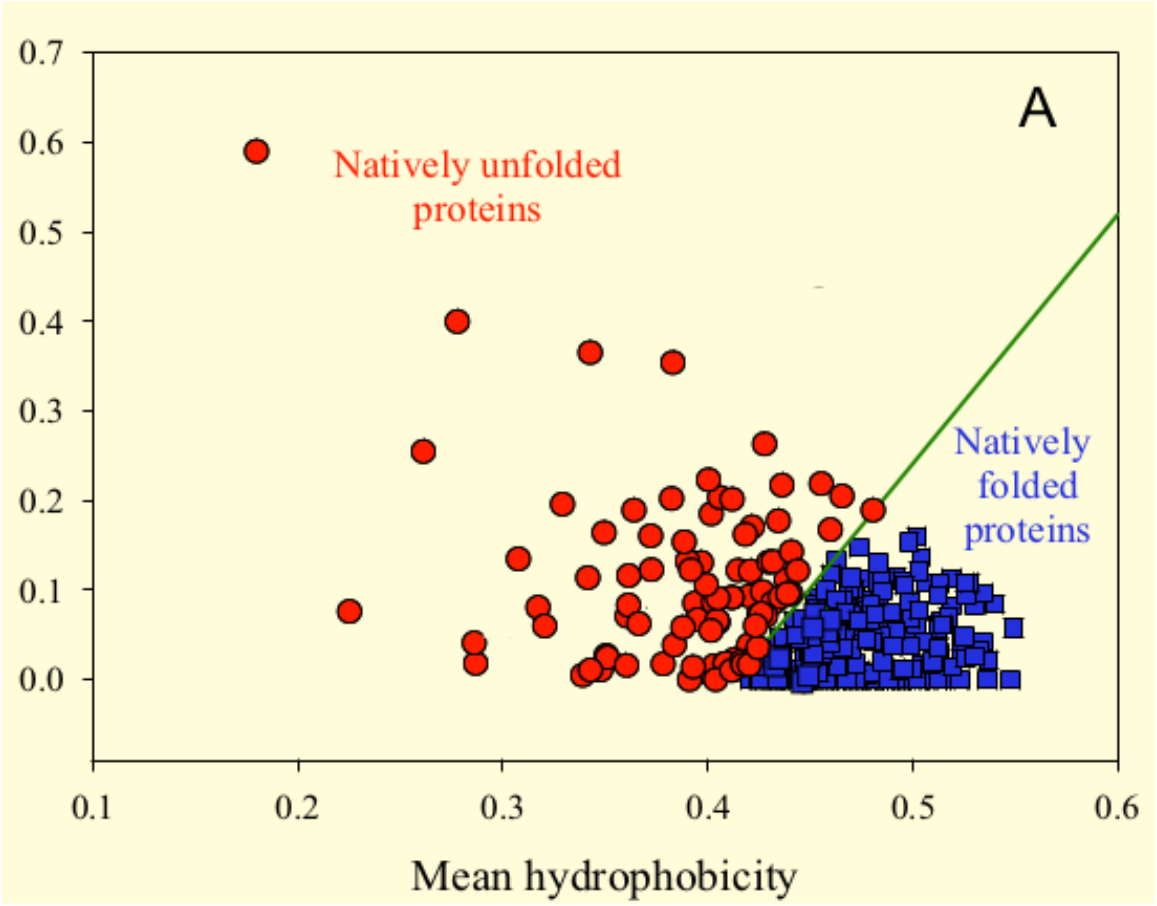


FIGURE 1.4: Charge-Hydropathy plot between unfolded and folded proteins [5]

IDPs are rich in so-called disorder-promoting aminoacids, which are polar and charged aminoacids (Metionine, Lysine, Arginine, Serine, Asparagine and Glutamic Acid) or structure breaking aminoacid (Proline and Glycine). On the other hand, IDPs show a significantly lack of order-promoting amino acid, which have an aliphatic or aromatic group (Valine, Leucine, Isoleucine, Tyrosine, Phenylanine, Tryptophane and Cysteine) [6, 20, 21]. These features are shown in Fig. 1.5.

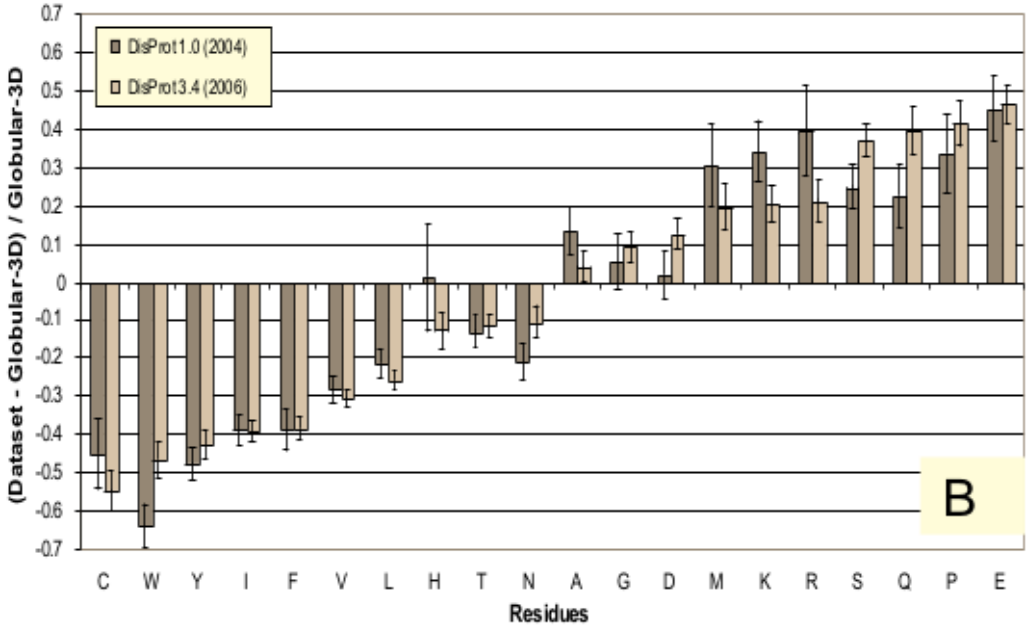


FIGURE 1.5: Amino acid composition of two sets of IDPs (Disprot 1.0 and Disprot 3.4), relative to a set of globular proteins (PDB 3D) [5]

It is known that proteins should be studied taking into account the physical-chemical environment in which they have to perform their function, since protein interaction with environment drives all biological processes. Being water present in every biological system, it is evident how the study of the interaction of proteins in water is fundamental for understanding their structure and function. The intrinsically disorder of this kind of

proteins makes the hydration of the IDPs much higher than the globular proteins, and there are different hydration states between totally and partially unfolded protein [20].

This physical properties make it possible the formation of different types of conformations that can cross the conformational space and occupy a particular state. IDPs cover a spectrum of states from fully unstructured to partially structured and include random coils, (pre) - molten globules and large multi - domain proteins connected by flexible linkers defined as continuum model (see Fig. 1.6).

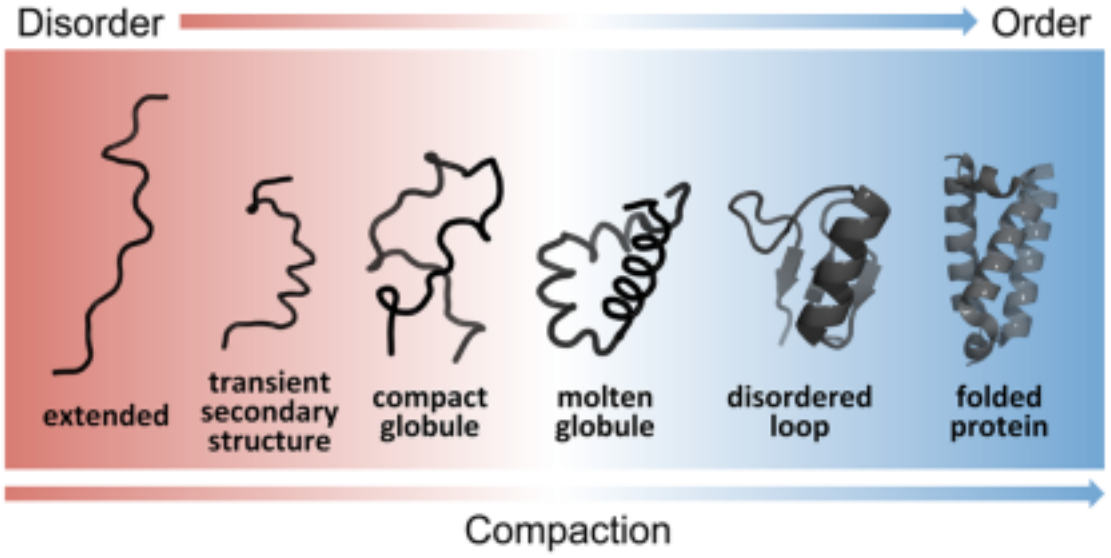


FIGURE 1.6: Continuum model of protein structure [6]

In this framework, there are no boundaries between the described states: native proteins could appear anywhere within this continuous landscape. IDPs domains are dynamic and fluctuate over an ensemble of heterogeneous conformations. These fluctuations are determined by the nature of the aminoacids and their distribution in the primary structure. This is why IDPs are studied as dynamic ensembles of rapidly interconverting conformations which can be described by relatively flat energy landscapes [6]. The characteristics

of these proteins give them many advantages over ordered proteins:

- great interaction surface;
- conformational flexibility;
- post - translational modifications to adjust their function in the cell;
- low affinity in interaction.

The lack of a well defined structure is the crucial point that gives a huge functional advantage to the IDPs, conferring functional characteristics absent in the globular proteins. For example, when the disorder parts play the role of flexible linkers between two rigid domains, these allow the rigid domains to be able to adapt to more substrate types, in a specific way. Moreover, when binding to the target site occurs, there is a strong decrease in conformational entropy and this unfavorable term separates the binding force from the specificity of action, making the process reversible, a fundamental characteristic in the processes of cell regulation. Another important feature is the high surface available in contact with other molecules that facilitates the interaction in several points causing an increase in the binding specificity and in binding velocity. IDPs are the unique proteins able to participate in both one-to-many and many-to-one signaling [20]. These characteristics allow proteins to perform more functions, in smaller sizes than globular proteins. Understanding these functions will be the object of study for many years [22].

1.1.2 IDPs Functions

Flexibility is the magic property that makes IDPs so predisposed to perform biological function. Studies have shown that IDPs are involved in the binding with nucleic acids, metal ions, protein-protein interactions and in the interaction with double lipid layers [20]. Based on their mode of action, they were classified into six functional groups [7, 22, 23]:

- Assemblers: proteins able to assembling, stabilizing multiprotein complexes such as the ribosome;
- Chaperones: molecules that can help protein's folding;
- Display sites: proteins able to molecular's recognition with regulatory and signaling function [24], which typically involve multiple partners and require high specificity and low affinity;
- Effectors: particular proteins that interact with lipid membranes, modifying their fluidity;
- Entropic chains: proteins that are able to modify their polypeptide chain to interfere with the position of other domains;
- Scavengers: proteins able to store or neutralize small lingands.

In this kind of classification is important that the various functional modes are not exclusive. A representative diagram of the functions performed by the IDPs is shown in the Fig. 1.7

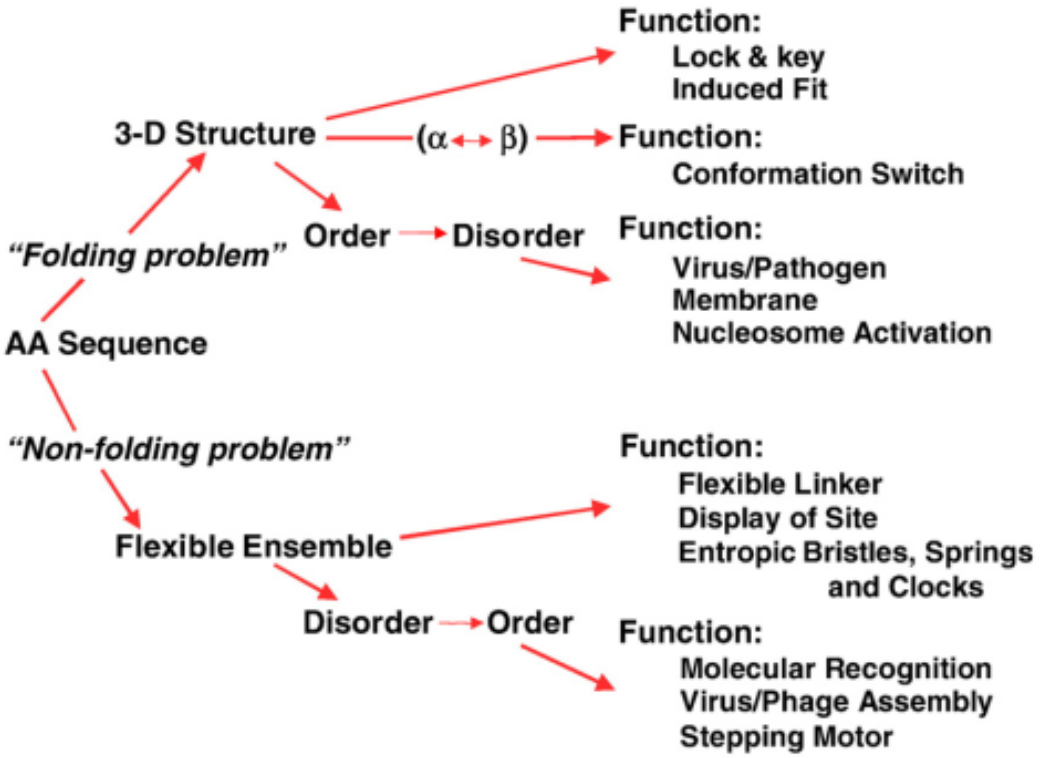


FIGURE 1.7: Involvement of intrinsic disordered in protein function [7]

IDPs that are able to perform their function undergo a folding caused by the interaction with other proteins, nucleic acids, membranes or small molecules. Folding can be caused by changes in the surrounding environment and the result of this process can lead to more or less compact structures.

1.1.3 D^2 concept: Disorder in diseases

Due of their important functions, IDPs are highly present within the human proteome and, although most of them are not pathological, some are implicated in human diseases [24]. Most IDPs, being described as dynamic ensembles, are indeed involved in different pathologies. The origin of the so-called D^2 concept, which indicates how the same protein can participate in a pathway of different diseases, originates from these features. For example, Tau protein is involved in the misfolding and aggregation processes leading to both Parkinson's and Alzheimer's diseases [20]. The most important diseases that involve IDPs are:

- Cardiovascular disease;
- Cancer;
- Neurodegenerative diseases;
- Amyloidoses.

This thesis deals with the investigation of α-syn, the IDP involved the Parkison's neurodegenerative disease. Hence, in the next paragraph the state of the art of α-syn will be reviewed.

1.1.4 α - synuclein: Neurodegerative's protein

α - synuclein has been discovered to be the main protein involved in one of the most important neurodegenerative diseases in the world: Parkinson's disease.

1.1.4.1 Parkinson's disease

Parkinson's disease is a progressive neurological disorder characterized by a large number of motor and non-motor characteristics that may have an impact on the patient's daily life. PD is the second most common neurodegenerative disease after Alzheimer's disease [8] and future predictions, as shown in the Fig. 1.8, are far from reassuring.

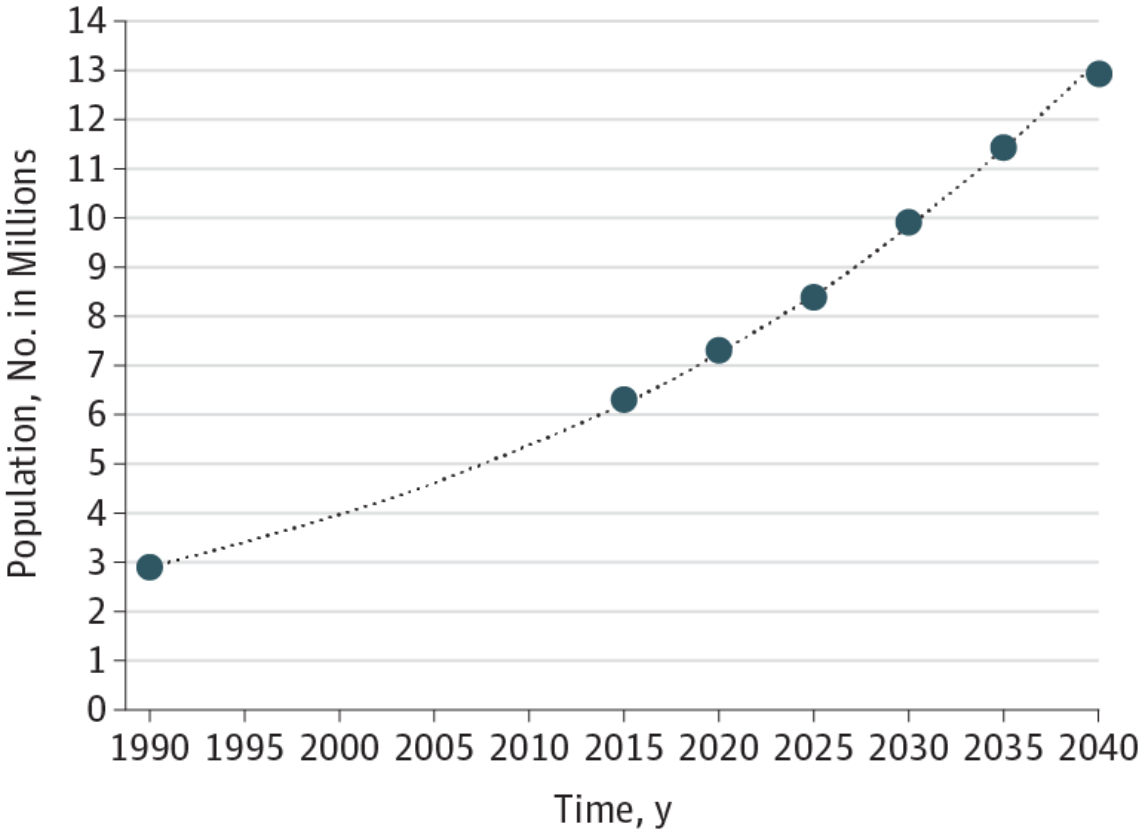


FIGURE 1.8: Estimated and projected number of individuals with PD, 1990-2040 [8]

The main important hallmarks of PD are the presence of cytoplasmatic inclusions called "Lewy bodies" and "Lewy neurities", and the selective degeneration and loss of dopaminergic neurons in the substantia nigra. The Lewy bodies and Lewy neurities are formed by filaments with length from 5 to 10 nm, known as amyloid fibers. The presence of

these fibers in the brain tissue is typical of neurodegenerative diseases and the process according to which they develop is shown in the Fig. 1.9.

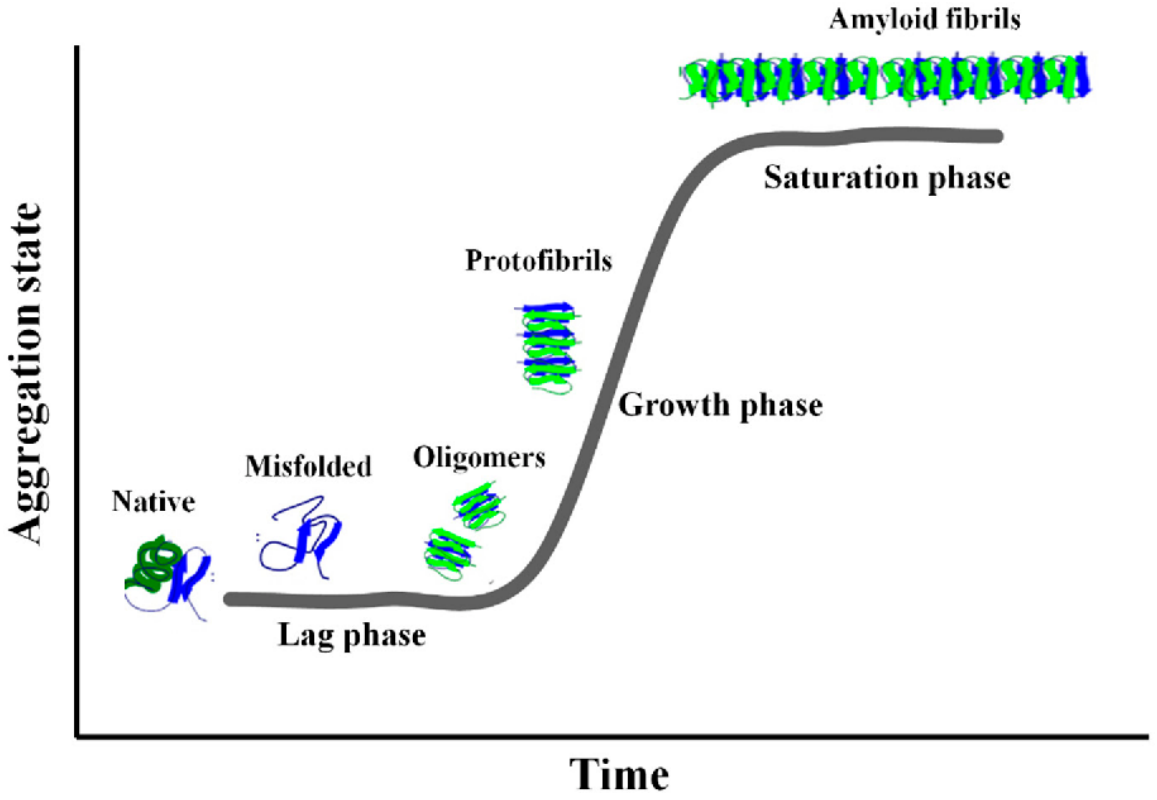


FIGURE 1.9: Aggregation process of IDPs [9]

This aggregation process, considered valid for all proteins that tend to form amyloid fibers, is described with a sigmoid curve divided into three phases:

- **Lag phase** in which the nucleation sites are created, which are the seeds for the development of future fibers;
- **Growth phase**, also called elongation phase. in which there is the formation of oligomers and protofibrils;
- **Saturation phase**, the final stage of the fibrillation process, in which there is the formation of mature amyloids fibrils.

PD is characterized by two main characteristics:

- the progressive damage of dopaminergic neurons that lead to the lack of the release of dopamine necessary for the maintenance of essential functions;
- the appearance of non-motor symptoms associated with degeneration of the non-dopaminergic system, due to the accumulation of Lewy's bodies in different brain regions.

The main motor symptoms of PD that, can be grouped under the acronym TRAP, are:

- Tremor at rest, the most easily recognised symptom of PD;
- Rigidity, characterised by increased resistance of the passive movement of a limb (e.g. flexion, extension or rotation movements);
- Akinesia, which refers to the slowness of movement and is the most characteristic clinical feature of PD;
- Postural instability, resulting in abnormal axial postures (e.g. scoliosis).

Furthermore, the flexed posture and the motor block were included among the classic features of parkinsonism, PD being the most common form [25].

The diagnosis of PD mainly relies on the clinical detection of these motor symptoms, but there is not a definitive diagnostic test. These symptoms are often accompanied by the appearance of autonomic, cognitive and psychiatric problems. PD, at the actual state, is an irreversible disease and an incurable disorder [10]: for all these reasons, there is a

great effort in trying to understand what is the molecular basis of the onset of the disease.

1.1.4.2 α - synuclein

α - synuclein is a IDP consisting of 140 amino acids, encoded by the SNCA gene, which is in unfolded state in water solution. α-syn is composed of three distinct regions [26]:

- Amino - terminus (aa 1-60);
- NAC (non- Aβ component, aa 60-95), which contains the hydrophobic region and confers the potential to from β-sheet secondary structure;
- Carboxyl terminus (aa 95-140), which is highly negatively charged. This tail is thought to drive protein- protein interactions, typical of the physiological functions of α-syn.

The scheme of this protein is show in Fig. 1.10

α-syn is present at high concentration, as it has been estimated representing up to 1% of the total proteins in the soluble cytosolic brain fractions. α-syn has several physiological functions:

- release and trafficking of synaptic vesicles;
- fatty acid bonds;
- physiological regulation of some enzymes;

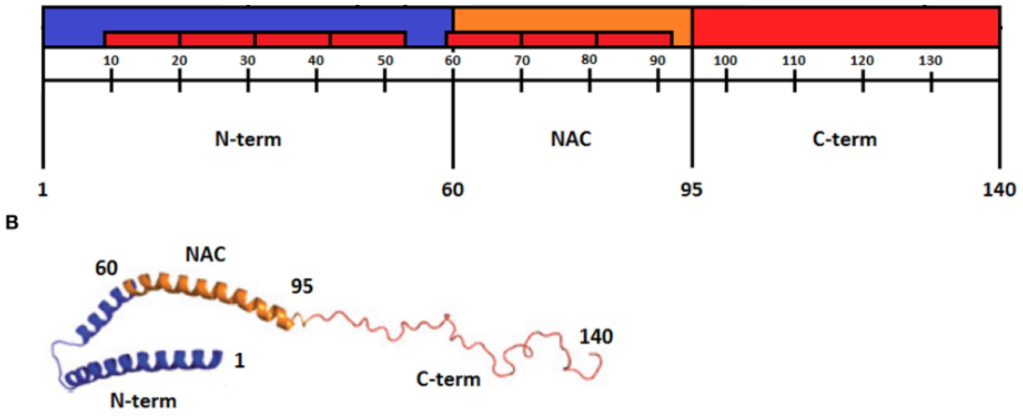


FIGURE 1.10: Scheme of α-syn sequence. Image modified from [10]

- transporters;
- neurotransmitter vesicles;
- roles in neuronal survival;
- important factors contributing to long-term functioning of the nervous system;
- cellular signaling [1].

α-syn is a natively unfolded protein in solution that shows a low hydrophobicity and high net charge, a common feature of most IDPs. Hydrophobicity and charge are characteristics that are easily modulated by the chemical-physical environment surrounding the protein, and therefore small changes in this environment regarding temperature or pH can change the conformation of this kind of protein [1].

The interest on the study of α-syn increased after the discovery of a gene mutation that was associated with the case of early - onset PD and its aggregates were found to be the major components of Lewy bodies [1].

1.1.4.3 α-synuclein mutants

During the studies carried out on Parkinson's disease, various missense mutations have been found, which have been associated with the case of early - onset PD [11]. The mutants analyzed in this thesis are:

- A53T;
- A30P;
- E46K;
- G51D.

Fig. 1.11 shows the comparison between the wild - type (WT) primary sequence and that of the mutants taken into consideration in this work, in order to highlight the position of the point mutation. All mutations considered shown in Fig. 1.11 are in the N - terminus and therefore in the figure only that part of the sequence has been reported.

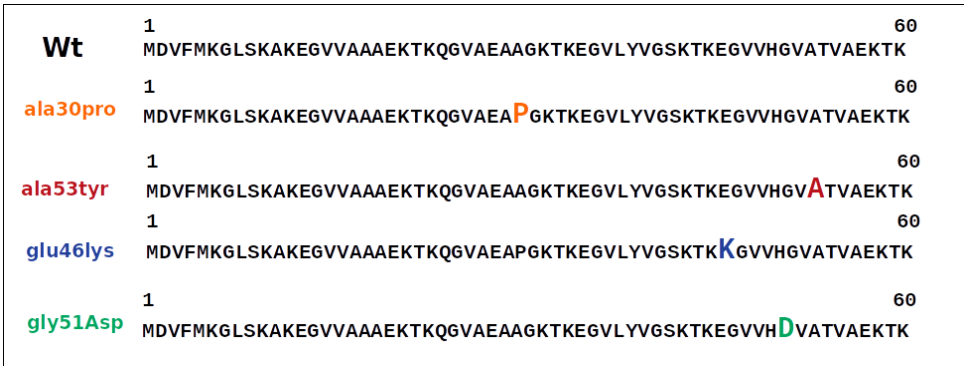


FIGURE 1.11: Primary sequence of α-synuclein 's mutants in comparison with WT sequence .

In Fig. 1.12 the substituted aminoacid (on top) and a schematic representation of the mutation in the sequence (on bottom) are shown.

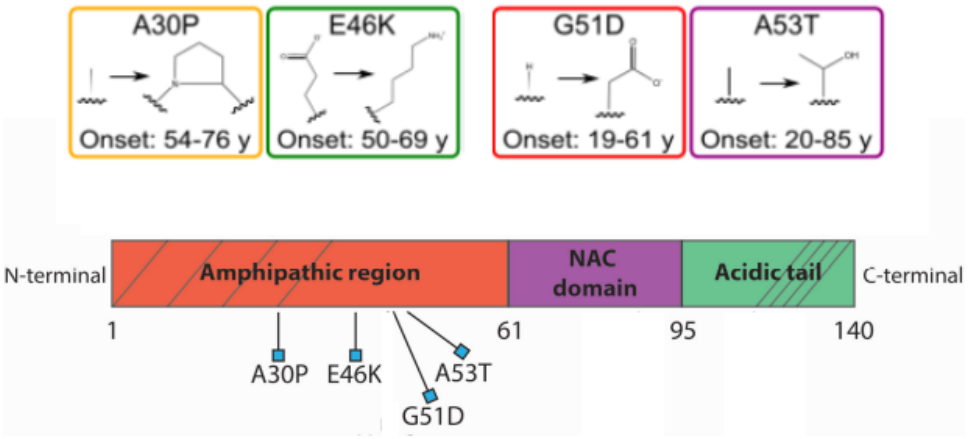


FIGURE 1.12: Top: substituted aminoacid in the point mutation [11]. Bottom: schematic representation of α-syn sequence with the point mutation. Bottom figure has been modified from Ref. [12].

The first effect caused by the mutations is the difference age in which the disease appears. The cases examined in this thesis are always faced with an early Parkinson disease, but in the literature also mutations that cause a delay in the appearance of symptoms are reported. The effects of these point mutations on the peptide aggregation kinetics have been widely studied. The A53T variant is considered to accelerate the aggregation, compared with the WT protein. The effects of A30P and G51D are instead under discussion, as it is reported both an increase and a decrease in the aggregation rate, while the effect of E46K is similar to that of the A53T but with a lower difference compared to the WT protein [11, 27].

By analyzing the A53T and A30P mutants with respect to the WT, it has been possible to evaluate their effect as a function of:

- Temperature;

- Concentration of the protein;
- β propensity.

The increasing of temperature, within a certain limit, causes an increase of aggregate structures. In fact, at low temperature all three types of proteins remain unfolded, while as the temperature increases, the mutants undergo a greater structural modifications compared to what happens in the WT protein. As far as the concentration effect is concerned, it is evident from literature that as the concentration of protein increases in solution, there is an increase in aggregation propensity. This increase is not the same for each mutant. The A53T mutant increases its propensity to aggregate more than the A30T, which is however more inclined to from aggregate structure than the WT [27]. Looking at the content of β structures, FTIR measurements, shown in [27], show that WT-α-syn results in an unfolded situation, while the A30P and A53T mutants, especially the latter one, show a greater presence of β structures.

The E46K mutant shows electrostatic properties similar to the WT protein, but exhibits, according to NMR measurements, conformational changes in the vicinity of the mutated amino acid, while the remaining part of the sequence does not present large conformational changes. This mutant shows a greater capacity, compared to the WT, in the formation of fibers, with a behavior very similar to the on of A53T mutant[28]. The G51D mutation doesn't affect the structure of free α-syn in solution and shows a slow aggregation rate in comparison with WT protein.

Chapter 2

Small Angle Scattering

2.1 SAS: An introduction to the technique

The experimental technique applied in this thesis is Small Angle Scattering (SAS), which provide information on the structure of particles that have typical dimensions between Angstrom (Å) and Nanometers (nm). Depending on the probe used, X-Ray or Neutron, this technique reveals differences in the electron scattering length density (SLD), proportional to the electron density, or in the neutron scattering length density, respectively. The IDPs experimental investigations carried out in this thesis have used the the synchrotron's X-Rays (Fig. 2.1), using the Small Angle X-Ray Scattering's technique (SAXS).

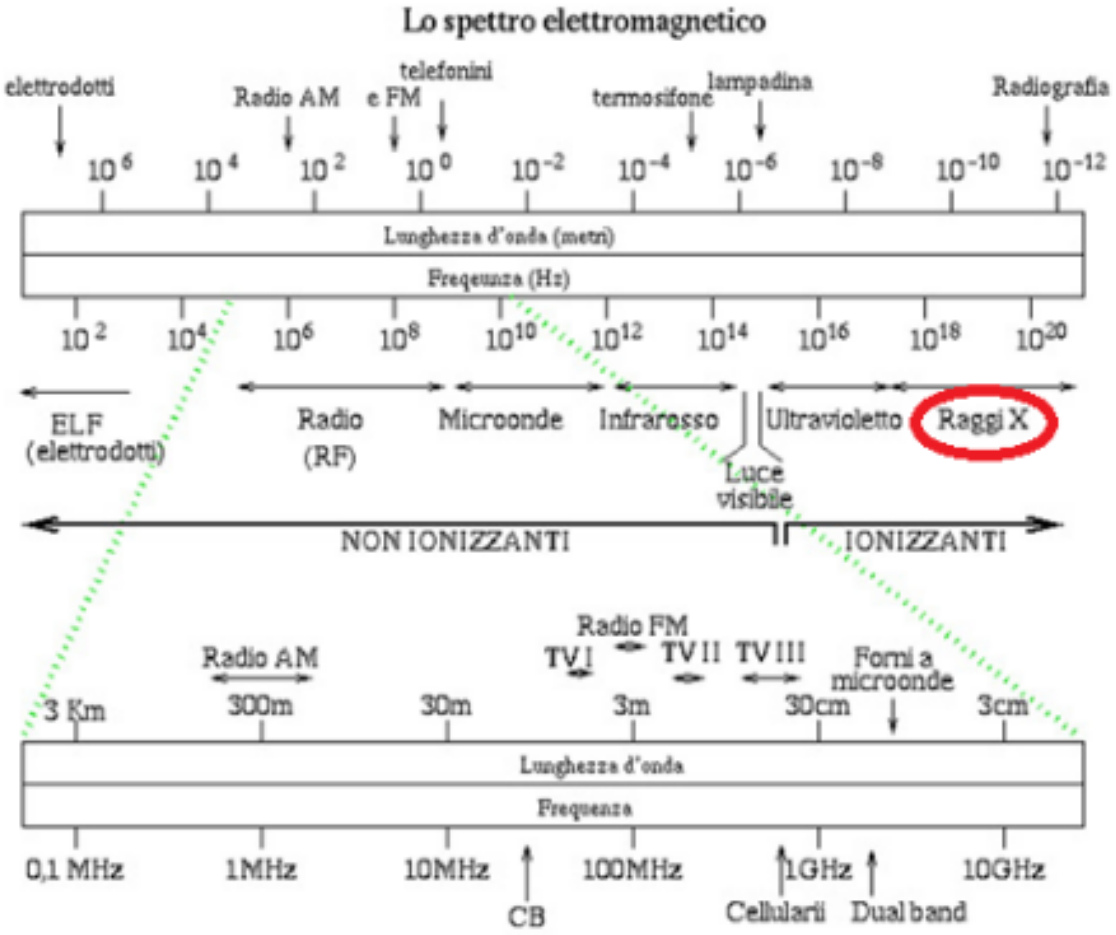


FIGURE 2.1: Electromagnetic spectrum

In Fig. 2.2 the resolution of the SAS techniques is compared with other experimental techniques. It should be noted that while SAXS and SANS are investigating the same order of magnitude, they are sensitive to different characteristics of the sample.

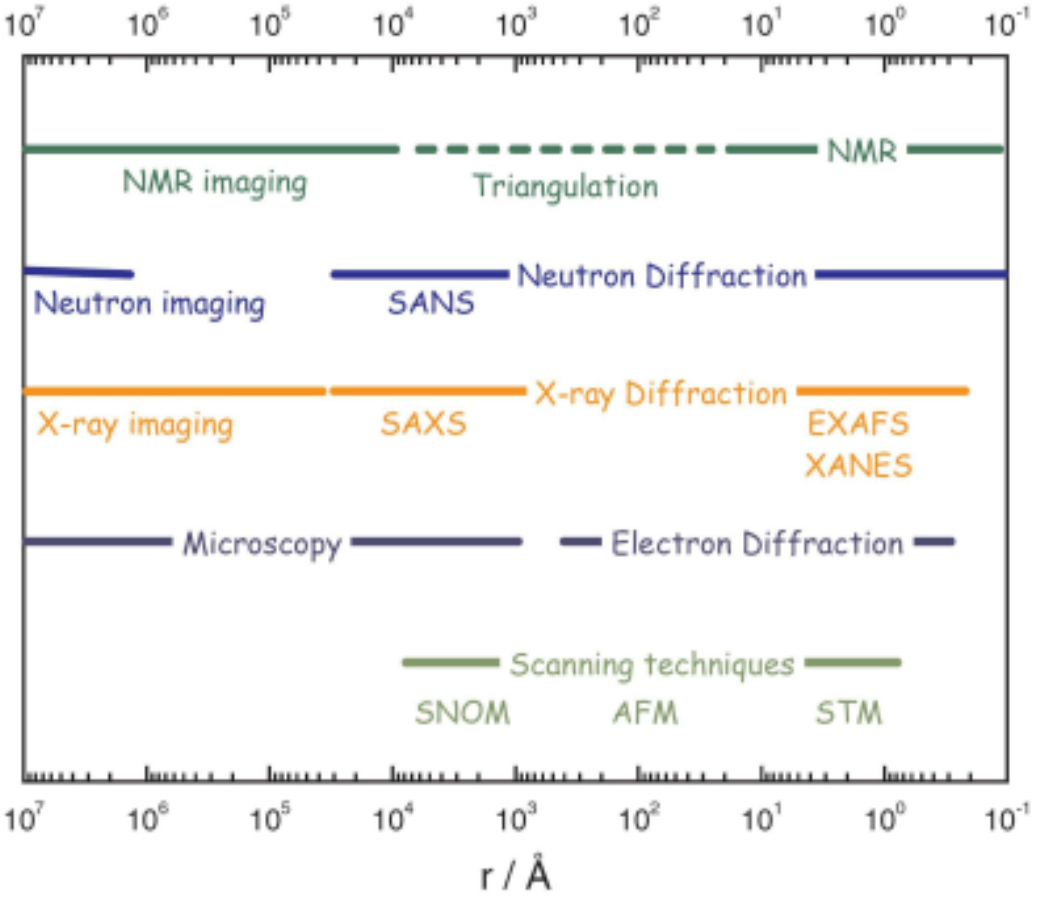


FIGURE 2.2: Resolution of SAS technique

The application of SAS in Biology is very useful because with this technique is possible to investigate samples directly in solution. In the study of protein, SAS can provide information about:

- quaternary structure;
- folded or unfolded status;

- conformational changes during the various active phases of the molecule;
- unfolding processes due to chemical or temperature treatments under pressure;
- interactions between different macromolecules in solution;
- hydration properties of macromolecules;
- kinetic process of protein aggregation.

The great advantage of SAS technique compared to others is that of being able to carry out the measurements in experimental conditions very similar to the *in vivo* environment and therefore to be able to study the macromolecules under different conditions of pressure, pH and temperature.

2.2 Small Angle Scattering geometry

The geometry of a SAS experiment is shown in Fig. 2.3. A beam of X-Rays or Neutrons impacts on the sample and this interaction generates the scattering phenomenon which is detected by a detector.

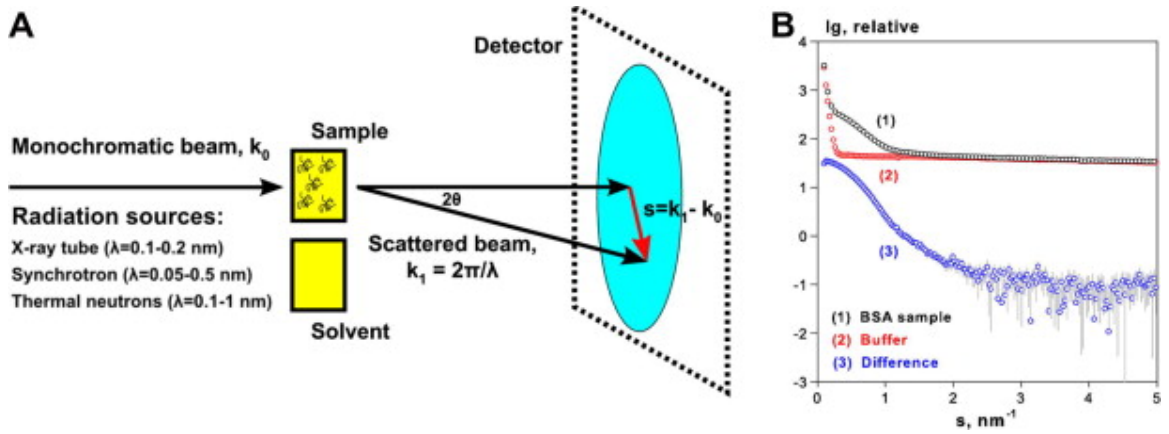


FIGURE 2.3: Schematic SAS experiment: A. Instrumental setup; B. Curve example. [13]

From a physical point of view, a SAS experiment can be schematized as in Fig. 2.4.

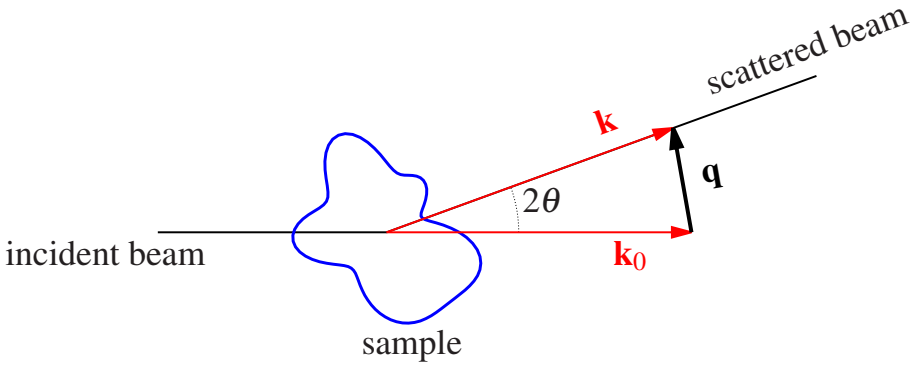


FIGURE 2.4: Representation of the scattering vector.

The scattering process is considered elastic in first approximation. As shown in Fig. 2.4, $\mathbf{k_0}$ and $\mathbf{k}$ are the incident and diffuse wave vectors respectively while 2θ is the scattering angle between the incident and the diffused vector.

The scattering vector $\mathbf{q}$ is defined as:

$$\mathbf{q} = (\mathbf{k} - \mathbf{k_0}) \tag{2.1}$$

and his modulus results:

$$q = \frac{4\pi}{\lambda} \sin\theta \tag{2.2}$$

where λ is the wavelength of the incident radiation and 2θ is the scattering angle.

When the wavelength λ has been selected, for any 2θ angle at which the scattered radiation is detected we can associate its scattering wave vector $\mathbf{q}$. From the Bragg's law for X-ray Diffraction (XRD) of a crystal from a characteristic distance d between two of its planes we have $n\lambda = 2d\sin\theta$ and substituting in the 2.2 is possible to deduce that in order to obtain information from the sample under study it is necessary to take into account the minimum distance that can be detected that is: $d_{min} \approx 2\pi/q_{max}$, q_{max} being the maximum value of the scattering vector modulus.

Hence, in order to carry on a proper SAS experiment, it is necessary to roughly know the dimensions of the objects to be investigated, in order to appropriately choose the

distance between the detector and the sample that will determine the appropriate range of the modulus of the scattering vector (q-range).

2.3 SAXS Theory

In this section some basic concepts necessary for the use of this technique will be explained.

2.3.1 Macroscopic differential cross section

The SAXS technique makes it possible to measure the so-called macroscopic differential scattering cross section, $\frac{d\Sigma}{d\Omega}(\mathbf{q})$, defined by the master equation:

$$\frac{d\Sigma}{d\Omega}(\mathbf{q}) = \frac{1}{V}\left\langle \left| \int_V d\mathbf{r}\ \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} \right|^2 \right\rangle \tag{2.3}$$

where V is the volume of the irradiated sample and the average is made on all possible positions and orientations of the particles in the system. Eq. 2.3 represents the three-dimensional Fourier transform (FT) of the electron SLD, $\rho(\mathbf{r})$, in the case of SAXS experiments, while for SANS $\rho(\mathbf{r})$ indicates the neutron SLD. The scattering length density $\rho(\mathbf{r})$ can be expressed as:

$$\rho(\mathbf{r}) = \delta\rho(\mathbf{r}) + \rho_0, \tag{2.4}$$

where ρ_0 is the average scattering density calculated on a larger volume than that resolution volume. The fundamental equation therefore becomes:

$$\begin{aligned}\frac{d\Sigma}{d\Omega}(\mathbf{q}) &= \frac{1}{V}\left\langle\left|\int_V d\mathbf{r}\ [\delta\rho(\mathbf{r})+\rho_0]e^{i\mathbf{q}\cdot\mathbf{r}}\right|^2\right\rangle \\ &= \frac{1}{V}\left\langle\left|\int_V d\mathbf{r}\ \delta\rho(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}+\rho_0\int_V d\mathbf{r}\ e^{i\mathbf{q}\cdot\mathbf{r}}\right|^2\right\rangle \end{aligned} \tag{2.5}$$

As $\frac{1}{(2\pi)^3}\int d\mathbf{r}\ e^{i\mathbf{q}\cdot\mathbf{r}}$ it is a representation of the three-dimensional Dirac function, we will have:

$$\frac{d\Sigma}{d\Omega}(\mathbf{q}) = \frac{1}{V}\left\langle\left|\int_V d\mathbf{r}\ \delta\rho(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}+\rho_0(2\pi)^3\delta(\mathbf{q})\right|^2\right\rangle \tag{2.6}$$

The contribute of the Dirac Delta is different from zero only for $\mathbf{q}=0$, therefore this term is not involved in a scattering experiment ($\mathbf{q}=0$ means that there was no diffusion). To formalize the presence of various diffusion centers within the sample and make explicit the term $\delta\rho(\mathbf{r})$, it is useful to define the dimensionless *scattering amplitude* of a particle:

$$F_i(\mathbf{q}) = \frac{1}{f_i}\int_{V_i} d\mathbf{r}\,\delta\rho_i(\mathbf{r})\,e^{i\mathbf{q}\cdot\mathbf{r}} \tag{2.7}$$

where V_i is the volume of i-th particle that diffuses and f_i is a normalization coefficient. Indeed, when $\mathbf{q}=0$, we have:

$$f_i = \int_{V_i} d\mathbf{r}\, \delta\rho_i(\mathbf{r}), \tag{2.8}$$

Considering a system constituted by N_p particles, with their center of mass in the vectors position $\mathbf{R}_i$ and with orientation ω_i with respect to the reference, each point k of the particle will be identified by the vector $\mathbf{r}_k = \mathbf{R}_i + \mathbf{u}_k$ (Fig. 2.5).

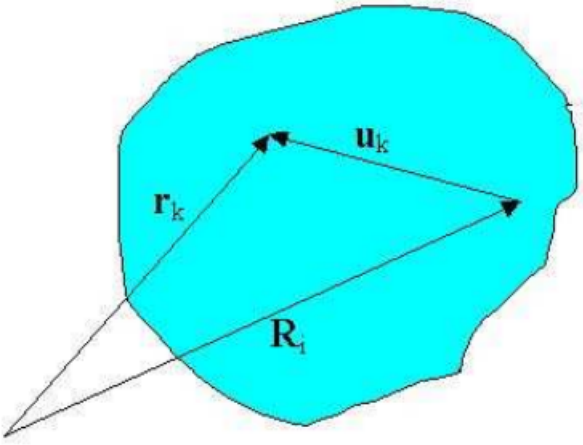


FIGURE 2.5: Representation of the position vector $\mathbf{R}_i$.

The equation 2.3 becomes:

$$\begin{aligned}
\frac{d\Sigma}{d\Omega}(\mathbf{q}) &= \frac{1}{V}\left\langle \sum_{i=1}^{N_P}\sum_{j=1}^{N_P}\int_{V_i}\int_{V_j} d\mathbf{u}_i d\mathbf{u}_j \delta\rho_i(\mathbf{u}_i)\delta\rho_j(\mathbf{u}_j) e^{i\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}_j)} e^{i\mathbf{q}\cdot(\mathbf{u}_i-\mathbf{u}_j)} \right\rangle \\
&= \frac{1}{V}\sum_{i=1}^{N_P} f_i^2 < F_i^2(\mathbf{q}) >_{\omega_i} \\
&\quad + \frac{1}{V}\sum_{i=1}^{N_P}\sum_{j\neq i}^{N_P} f_i f_j < F_i(\mathbf{q}) F_j^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}_j)} >_{\omega_i,\omega_j,\mathbf{R}_i,\mathbf{R}_j},
\end{aligned} \tag{2.9}$$

where the first average is done on all possible orientations ω_i of the particle i-th and the second average is made on the combinations of orientations ω_i e ω_j and on the positions $\mathbf{R}_i$ and $\mathbf{R}_j$ of the particles i and j. For the calculation of these averages it is useful to introduce the distribution function related to the single particle $P^{(1)}(\omega_i)$ and the torque distribution function $P^{(2)}(\mathbf{R}_i, \omega_i, \mathbf{R}_j, \omega_j)$ such that:

$$< F_i^2(\mathbf{q}) >_{\omega_i} = \frac{\int d\omega_i \, P^{(1)}(\omega_i) \, F_i^2(\mathbf{q})}{\int d\omega_i \, P^{(1)}(\omega_i)} \tag{2.10}$$

$$< F_i(\mathbf{q}) F_j^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_i - \mathbf{R}_j)} >_{\omega_i, \omega_j, \mathbf{R}_i, \mathbf{R}_j} = \frac{\int d\omega_i d\omega_j d\mathbf{R}_i d\mathbf{R}_j \, P^{(2)}(\mathbf{R}_i, \omega_i, \mathbf{R}_j, \omega_j) \, F_i(\mathbf{q}) F_j^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_i - \mathbf{R}_j)}}{\int d\omega_i d\omega_j d\mathbf{R}_i d\mathbf{R}_j \, P^{(2)}(\mathbf{R}_i, \omega_i, \mathbf{R}_j, \omega_j)} \tag{2.11}$$

The first term of the equation 2.9 describes the shape of the particle, while the second term involves both the form and the correlation with other particles. This is why we define the average *form factor* the following term:

$$P(\mathbf{q}) = \frac{\sum_{i=1}^{N_P} f_i^2 < F_i^2(\mathbf{q}) >_{\omega_i}}{\sum_{i=1}^{N_P} f_i^2} \tag{2.12}$$

and the *structure factor* this term:

$$S(\mathbf{q}) = 1 + \frac{\sum_{i=1}^{N_P} \sum_{j \neq i}^{N_P} f_i f_j < F_i(\mathbf{q}) F_j^*(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_i - \mathbf{R}_j)} >_{\omega_i, \omega_j, \mathbf{R}_i, \mathbf{R}_j}}{sum_{i=1}^{N_P} f_i^2 < F_i^2(\mathbf{q}) >_{\omega_i}} \tag{2.13}$$

Obtaining:

$$\frac{d\Sigma}{d\Omega}(\mathbf{q}) = n_P < (\Delta\rho V_P)^2 > P(\mathbf{q})S(\mathbf{q}), \tag{2.14}$$

where n_p is the average particle number density, N_P/V, and $< (\Delta\rho V_P)^2 >= \frac{1}{N_P}\sum_{i=1}^{N_P} f_i^2$.

In this way we obtain a simple expression (2.14) that separates information about the shape of the objects investigated and their correlations.

2.3.2 Two - states model

In most SAXS experiments on proteins we can consider each particle as an object with uniform SLD, ρ_i, embedded in a solvent of homogeneous SLD ρ_0. Working in very diluted conditions, we can assume that ρ_0 is the average SLD of the system. Hence in Eqs. 2.7 and 2.8 the function $\delta\rho_i(\mathbf{r})$ can be replaced by the constant $\Delta\rho_i = \rho_i - \rho_0$:

$$F_i(\mathbf{q}) = \frac{1}{V_i}\int_{V_i} d\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}} \tag{2.15}$$

$$f_i = V_i\Delta\rho_i. \tag{2.16}$$

It can be shown that in the case of spherical particles with a uniform SLD, the expression of diffuse intensity becomes:

$$\frac{d\Sigma}{d\Omega}(q) = \sum_{\alpha,\beta=1}^{p} \sqrt{n_\alpha n_\beta}\Delta\rho_\alpha\Delta\rho_\beta V_\alpha V_\beta \Phi(qR_\alpha)\Phi(qR_\beta)S_{\alpha\beta}(q), \tag{2.17}$$

where the form factor for a sphere of radius R is:

$$F(q,R) = \Phi(qR) = 3\,\frac{\sin(qR) - (qR)\cos(qR)}{(qR)^3}. \tag{2.18}$$

The *structure factor* is:

$$S(q) = \frac{\sum_{\alpha,\beta=1}^{p} \sqrt{n_\alpha n_\beta}\Delta\rho_\alpha\Delta\rho_\beta V_\alpha V_\beta \Phi(qR_\alpha)\Phi(qR_\beta)S_{\alpha\beta}(q)}{\sum_{\alpha=1}^{p} n_\alpha(\Delta\rho_\alpha)^2 V_\alpha^2 \Phi^2(qR_\alpha)}. \tag{2.19}$$

When we are dealing with even more diluted system ($n_\alpha \to 0$),then we can approximate the structure factor $S_{\alpha\beta}(q)$ with the Kronecker delta function $\delta_{\alpha,\beta}$ as we suppose that the particles diffuse the radiation independently from each other. In the simplest case of a sample with a very diluted, monodisperse, two-phase and isotropic approximation, the diffuse scattering intensity reduces to:

$$\frac{d\Sigma}{d\Omega}(q) = n_P(\Delta\rho)^2 V_P^2 P(q), \tag{2.20}$$

where V_p is the single particle's volume.

These approximations simply indicates that with the SAS technique it is possible to obtain information about the shape of the investigated object. Assuming different shapes, corresponding form factors can be calculated. Some examples are shown in Fig. 2.6.

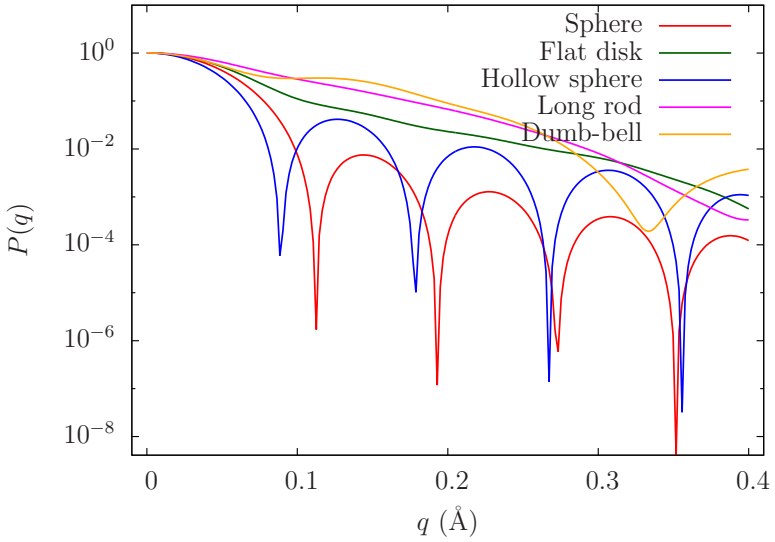


FIGURE 2.6: Examples of $P(q)$ of particles with different shape [14]

2.4 Data analysis: fundamental approaches

In the following paragraphs some fundamental approaches in the SAXS data analysis will be explained.

2.4.1 Guinier approximation

It is possible to obtain information about the average size of a particle investigated by SAS by exploiting the *Guinier approximation*. This approximation starts from the

development of the average form factor according to the hypothesis of randomly oriented particles:

$$\begin{aligned} <F^2(\mathbf{q})>_{\omega_q} &= \frac{1}{4\pi}\int d\omega_q \,\frac{1}{f^2}\int_{V_P}\int_{V_P} d\mathbf{r}_1 d\mathbf{r}_2 \,\delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_2)\, e^{i\mathbf{q}\cdot(\mathbf{r}_2-\mathbf{r}_1)} \\ &= \frac{1}{f^2}\int_{V_P}\int_{V_P} d\mathbf{r}_1 d\mathbf{r}_2 \,\delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_2)\frac{1}{4\pi}\int d\omega_q \, e^{i\mathbf{q}\cdot(\mathbf{r}_2-\mathbf{r}_1)} \\ &= \frac{1}{f^2}\int_{V_P}\int_{V_P} d\mathbf{r}_1 d\mathbf{r}_2 \,\delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_2)\frac{\sin\left(q|\mathbf{r}_2-\mathbf{r}_1|\right)}{q|\mathbf{r}_2-\mathbf{r}_1|}. \end{aligned} \tag{2.21}$$

Developing $\frac{\sin x}{x}$ in a truncated Taylor series up to the second order and considering the origin of the coordinate axes in the particle's SLD center in such a way that $\int_{V_P} d\mathbf{r}\,\delta\rho(\mathbf{r})\, r_\alpha = 0$ where $\alpha = x, y, z$, we obtain:

$$<F^2(\mathbf{q})>_{\omega_q} = 1 - \frac{q^2}{3f}\int_{V_P} d\mathbf{r}\, r^2\,\delta\rho(\mathbf{r}) + \ldots \tag{2.22}$$

At this point we can define in analogy with classical mechanics the *gyration radius* (Rg) of the particle respect to its center of SLD mass:

$$\begin{aligned} R_g^2 &= \frac{1}{f}\int_{V_P} d\mathbf{r}\, r^2\,\delta\rho(\mathbf{r}) \\ &= \frac{\int_{V_P} d\mathbf{r}\, r^2\,\delta\rho(\mathbf{r})}{\int_{V_P} d\mathbf{r}\,\delta\rho(\mathbf{r})} \end{aligned} \tag{2.23}$$

and then:

$$< F^2(\mathbf{q}) >_{\omega_q} = 1 - \frac{q^2 R_g^2}{3} + \dots \tag{2.24}$$

$$\approx \exp\left(-\frac{q^2 R_g^2}{3}\right) \tag{2.25}$$

This expression is known as *Guiner's law* and coincides with the form factor for terms up to the fourth power of q. This relationship loses validity with an increase of q. It can be shown that the Guinier's law is valid if the range of the exchanged wave vector is such that $q\,R_g < 1 - 1.3$. The gyration radius of a particle can be directly associated to the geometric characteristics of the sample. For example, for a sphere of radius R, we have that $R_g^2 = \frac{5}{3}R^2$; for an ellipsoid of semi-axes A, B and C we have ${R_g}^2 = (A^2 + B^2 + C^2)/5$. For other regular geometrical shapes, it is possible to determine a relationship between the gyration radius and the characteristic geometrical parameters of the i-th particle.

If in solution there are N_P species of particles, the form factor becomes:

$$P(q) = \frac{\sum_{i=1}^{N_P} f_i^2 \left(1 - \frac{q^2 R_{g,i}^2}{3}\right)}{\sum_{i=1}^{N_P} f_i^2} + \dots = 1 - \frac{q^2 R_{g,a}^2}{3} + \dots$$

$$\approx \exp\left(-\frac{q^2 R_{g,a}^2}{3}\right), \tag{2.26}$$

where the apparent squared gyration radius is weighted on the factors f_i^2:

$$R_{g,a}^2 = \frac{\sum_{i=1}^{N_P} f_i^2 R_{g,i}^2}{\sum_{i=1}^{N_P} f_i^2}. \tag{2.27}$$

In Fig. 2.7 an example of Guinier's plot is shown.

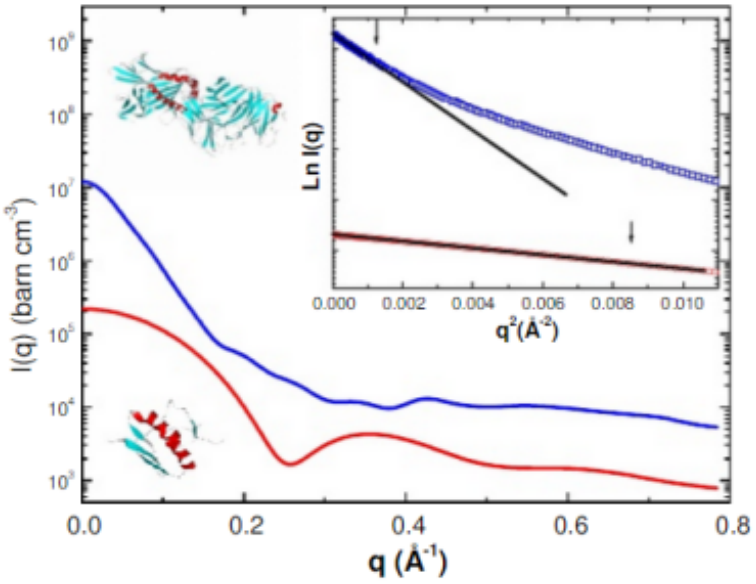


FIGURE 2.7: *Guinier's plot.* Theoretical SAXS curves calculated using the SASMOL software [15] for two different globular proteins. The box shows the corresponding plot made using the law of Guinier [14].

2.4.2 Kratky plot

Kratky's plot is a particular graphical representation of experimental data that provides information on the compactness of the investigated object. The Kratky plot shows the product $\frac{d\Sigma}{d\Omega}(q)\, q^2$ *vs* q. This representation is particularly useful for structural characterization of globular particles and for determining the different stages denaturing. In Fig. 2.8 an example of Kratky's plot from theoretical SAXS curve of different Bovinum Serum Albumin (BSA) conformations is shown. In blue the typical bell-shaped trend of a globular protein is shown, red curve reports the trend of a molten - globule state and finally in black the trend of an unfolded protein is shown.

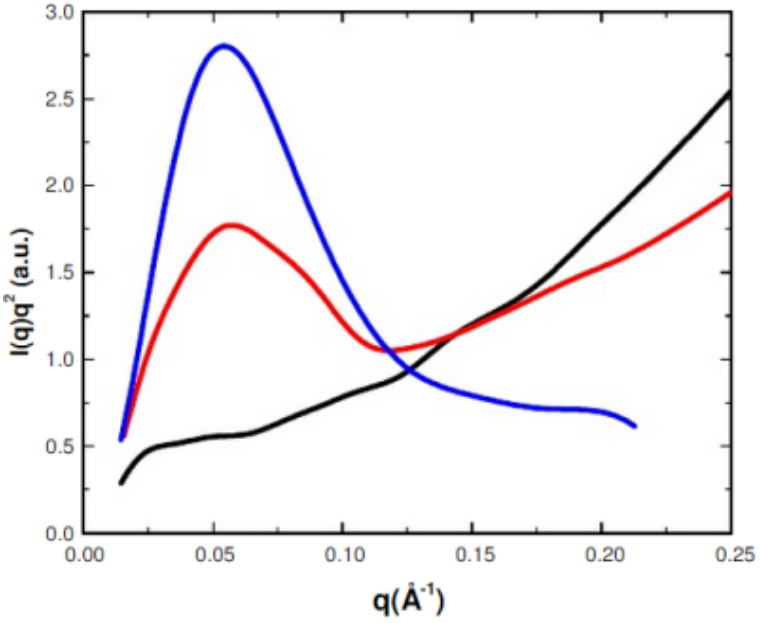


FIGURE 2.8: *Kratky's plot* of theoretical SAXS curve of different BSA conformation [14]

Because the peak position of the Kratky's plot is related to R_g of the objects in solution, a variation of the position of the peak reflects a change in particles size. In particular, the R_g increases if the peak of the Kratky's plot moves to lower q. If the width and the shape of the peak changes, it means that the object in solution is undergoing consistent structural changes.

2.4.3 Distance distribution function and Porod approximation

The *Porod's approximation*, unlike the Guinier's approximation, is useful for information on the volume and exposed surface of the particles in solution and is valid for large values of q. To derive the Porod's approximation, it is firstly necessary to start from the expression of the form factor of the particle:

$$< F^2(\mathbf{q}) >_{\omega_q} = P(q) = \int_0^\infty dr\, p(r)\, \frac{\sin(qr)}{qr}, \tag{2.28}$$

where $p(r)$ is defined as the inverse isotropic Fourier transforms of $P(q)$:

$$p(r) = \frac{2r}{\pi} \int_0^\infty dq\, P(q)\, q\, \sin(qr), \tag{2.29}$$

with the normalization condition $\int_0^\infty dr\, p(r) = 1$.

If we replace the equation 2.21 with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, we obtained:

$$\begin{aligned} P(q) &= \frac{1}{f^2} \int_{V_P} \int_{V_P} d\mathbf{r}_1 d\mathbf{r}\, \delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_1 + \mathbf{r}) \frac{\sin(qr)}{qr} \\ &= \int_0^\infty dr \frac{\sin(qr)}{qr} \left\{ 4\pi r^2 \frac{1}{f^2} \left\langle \int_{V_P} d\mathbf{r}_1\, \delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_1 + \mathbf{r}) \right\rangle_{\omega_r} \right\}. \end{aligned} \tag{2.30}$$

In relation to the equation 2.29 we deduce that we can express $p(r)$ as:

$$p(r) = 4\pi r^2 \frac{1}{f^2} \left\langle \int_{V_P} d\mathbf{r}_1\, \delta\rho(\mathbf{r}_1)\delta\rho(\mathbf{r}_1 + \mathbf{r}) \right\rangle_{\omega_r}, \tag{2.31}$$

showing that $p(r)$ is related to the autocorrelation function of $\delta\rho(\mathbf{r})$. It can be shown that the distribution function of distances $p(r)$ it can also be used to determine the R_g of the particles in solution:

$$R_g^2 = \frac{1}{2}\int_0^\infty dr\, r^2\, p(r). \tag{2.32}$$

For the two-phase model, the physical meaning of the $p(r)$ can be elucidated. Indeed, if we replace the function $\delta\rho(\mathbf{r})$ with the contrast product $\delta\rho$ (which we assume to be constant) and of the function $s(\mathbf{r})$ that describes the position and the shape of the particle ($s(\mathbf{r})$ is equal to 1 if $\mathbf{r}$ is inside the particle, 0 if it is outside the particle), we have:

$$p(r) = 4\pi r^2 \frac{1}{V_P^2}\int_{V_P} d\mathbf{r}_1 < s(\mathbf{r}_1+\mathbf{r}) >_{\omega_r} . \tag{2.33}$$

It is clear that the $p(r)$ represents the probability density to find a vector having the first extreme on the point $\mathbf{r}_1$ of the particle and the second extreme still in the particle and at a distance r from $\mathbf{r}_1$. If we develop the integral in the equation 2.33 considering separately the particle (*core*) and the volume of an inner shell of thick r (which we call *shell*), we obtain:

$$\begin{aligned}\int_{V_P} d\mathbf{r}_1 < s(\mathbf{r}_1+\mathbf{r}) >_{\omega_r} &= \int_{\text{core}} d\mathbf{r}_1 < s(\mathbf{r}_1+\mathbf{r}) >_{\omega_r} \\ &+ \int_{\text{shell}} d\mathbf{r}_1 < s(\mathbf{r}_1+\mathbf{r}) >_{\omega_r},\end{aligned} \tag{2.34}$$

The expression $<s(\mathbf{r}_1+\mathbf{r})>_{\omega_r}= \frac{1}{4\pi}\int d\omega_r s(\mathbf{r}_1+\mathbf{r})$ represents the fraction of a solid angle from which the point $\mathbf{r}_1$ can "see" the particle at a distance r. Hence, the argument of the integral on the *core* of the particle is always 1 so that the value of this integral is $V_P - S_P r$. For the integral on the inner *shell* we have:

$$\int_{\text{shell}} d\mathbf{r}_1 <s(\mathbf{r}_1+\mathbf{r})>_{\omega_r} \;=\; S_P \int_0^r dx \; <s(x+\mathbf{r})>_{\omega_r}, \tag{2.35}$$

where S_P is the particle's surface and x is depth's surface. Developing the integral on the solid angle:

$$\int_{V_P} d\mathbf{r}_1 <s(\mathbf{r}_1+\mathbf{r})>_{\omega_r} \;=\; V_P - S_P r + \frac{3}{4} S_P r = V_P(1 - \frac{S_P r}{4V_P}). \tag{2.36}$$

Then the $p(r)$ (2.33) becomes:

$$p(r) \;=\; \frac{4\pi r^2}{V_P}\left(1 - \frac{S_P r}{4V_P}\right) + \ldots \qquad \text{at small value of } r \tag{2.37}$$

which can be inserted into the expression of the form factor 2.29 to study its trend, at large values of q, as a function of r_{max} that is r value for which $p(r) = 0$:

$$P(q) = \frac{2\pi S_P}{q^4 V_P^2} + \frac{A(qr_{max})}{q^3 V_P^2}\sin(qr_{max}) + \frac{B(qr_{max})}{q^3 V_P^2}\cos(qr_{max}) + \ldots \qquad \text{at large value of } q. \tag{2.38}$$

So the trend of the $P(q)$ can be approximated by the *Porod's law*:

$$P(q) \approx \frac{2\pi S_P}{q^4 V_P^2}. \tag{2.39}$$

The approximation of Porod, with Guinier's approximation, allow us to carry out a simple but appropriate analysis on a SAXS curve. The use of both approximations is to date the method of data analysis of SAS more common in literature.

2.5 SAS from PDB structures

It is possible to calculate the form factor $P(\mathbf{q})$ starting from a structure deposited in the Protein Data Bank. Over the years, many methods have been developed to calculate the SAXS or the SANS profile from a PDB file [29, 30] and in this thesis work the SASMOL method was used. This method allows the determination of number and position of water molecules in the first three hydration shells [15] and is able to calculate the protein form factor by taking into account the contribution of the hydration shells which can

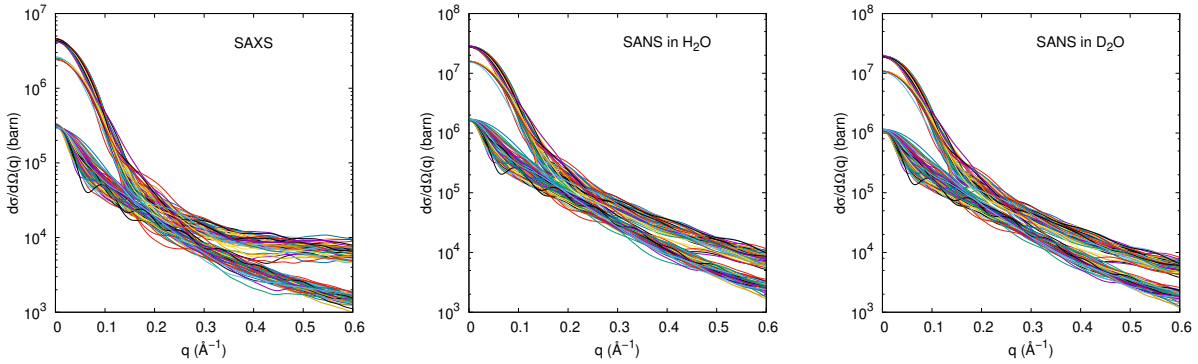


FIGURE 2.9: Simulated curves on the conformational ensemble used in this work (Section 4.3) obtained using the SASMOL method

have different scattering length densities, according to the cosolvent distribution in the local domain. In the figure 2.9 are shown some examples of the use of the SASMOL method, applied to the ensemble used in this thesis work. Simulated SAXS curves (left panel), simulated SANS curves in water (center panel) and simulated SANS curves in a complete deuteration situation (right panel) are reported. It is important to note that the SASMOL method can also simulate SANS curves at different deuteration ratios. As you can see from the figure the scattering intensities are different by changing the type of probe. The curves shown in the figure 2.9 were made starting from the pdb structure derived from [17], in which the ensemble consisted of three species (monomers, trimers and tetramers of α-syn). This subdivision is in the plot of the simulated curves. This subdivision is evident in the simulated curves in which there are three different trends.

2.6 SAS of IDPs

In the last twenty years, IDPs have been discovered to be fundamental molecules in a large number of biological processes, as shown in Chapter 1 and the importance of this

kind of protein has generated strong interest from researchers in order to understand the structural bases of their functions [16]. Nuclear magnetic resonance (NMR) was the most important structural biology technique used to characterize these structures at the level of the primary sequence and to discover the most disordered regions of the sequence[31]. Despite the importance of the NMR technique, some structural characteristics relating to the size and overall shape of the IDPs or their complexes remain invisible to NMR [16]. The best technique to obtain this information complementary to the NMR technique (Fig. 2.10) is SAS technique, of X-Rays or of Neutrons[32].

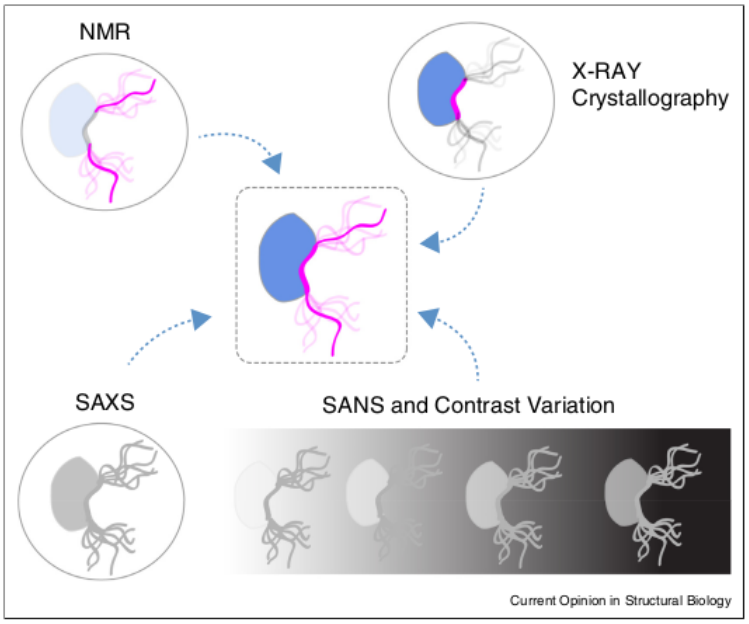


FIGURE 2.10: Representation of the structural sensitivity of NMR, X-ray crystallography, and SAS for a complex involving a IDP (central cartoon). NMR normally probes the flexible regions of these complexes while the globular partner and the interacting region remain invisible. Crystallography provides detailed information of the interacting region of the complex but not for the flexible parts. SAXS probes the complete ensemble, although the details cannot be assessed due to its inherent low-resolution. SANS, through contrast variation experiments, can probe independently both partners in the context of the complex depending on the deuteration level of the partners and the D_2O/H_2O of the buffer. SAS is an ideal tool to integrate NMR and crystallographic information to build complete structural and dynamic models of disordered biomolecular complexes [16].

Although SAS is a low-resolution technique, SAS curves do contain information on large-scale protein fluctuations and on the presence of multiple conformation in solution. The great challenge is to be able to interpret the SAS data in order to obtain useful structural information despite the enormous conformational variability. In order to correctly interpret the structural information contained in SAS data it is necessary to use three-dimensional models. However, the generation of conformational sets of disordered proteins is extremely difficult, mainly due to the flat energy landscape and, consequently, to a large number of local minima separated by low-energy barriers[16]. The most important

numerical methods for generating three-dimensional IDPs models are based on conformational landscapes derived from large databases of crystallographic structures [33]. The principal limitation of these approaches is the lack of consideration of long-range interactions between distant regions of the proteins. To describe these characteristics, force fields are needed that take into account the interactions within the chain and the features of the solvent. The development of specific force fields to study conformational fluctuations in IDPs is a very active field of research [34, 35]. Simulations of Molecular Dynamics (MD) or Monte-Carlo (MC), with a precise description of the energy of the system, are the best methods for correctly sampling the conformational space of these proteins [16]. As a matter of fact, the method developed in this thesis needs a conformational ensemble derived from computational results to interpret SAS data and be able to obtain, in this way, a description of the conformational behavior of the measured sample.

Chapter 3

Experimental Data

3.1 SAXS - Experimental Data

The experimental data reported in this thesis have been recorded at the beam line BM29 at the European Synchrotron (ESRF) in Grenoble (Fig. 3.1).

FIGURE 3.1: The European Synchrotron Radiation Facilities, Grenoble, France

The beam line BM29 (Fig. 3.2) is a tunable energy beamline (from 7 to 15 keV) for SAXS experiments of biological macromolecule solutions with the goal to determine their three-dimensional structures in a natural state with a "low" resolution (from nm to Å). The achievable q - range is $0.0025 - 0.6$ Å^{-1}, which corresponds to the biggest measurable Rg of the investigated particle of ~ 200 Å.

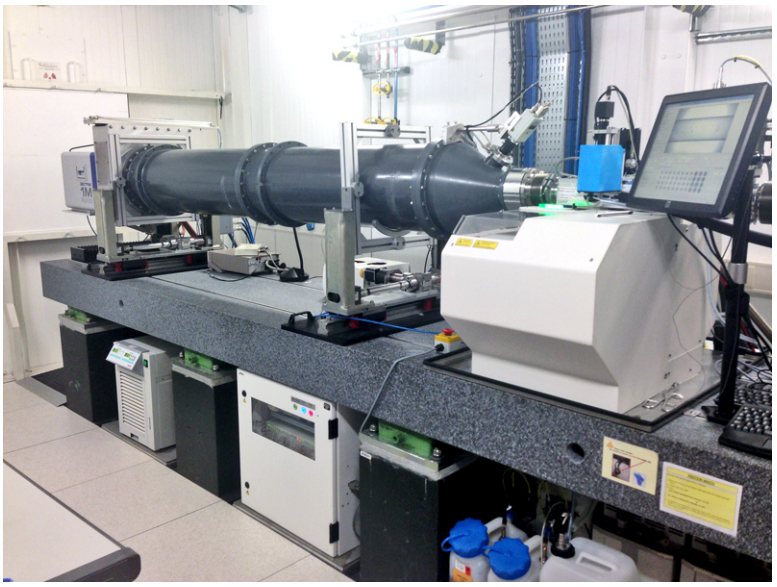


FIGURE 3.2: Beam Line 29 at The European Synchrotron Radiation Facilities, Grenoble, France

In the following paragraphs I will briefly illustrate the experimental sessions showing some SAXS curves recorded on α-synuclein samples.

3.2 α-synuclein

α-synuclein has been studied at different concentrations and at temperature of 25°, 37° and 45°C. Four different mutants have been investigated, in addition to the WT protein

in order to elucidate how a point mutation (Fig. 3.3) can modify the conformational landscape of this IDP.

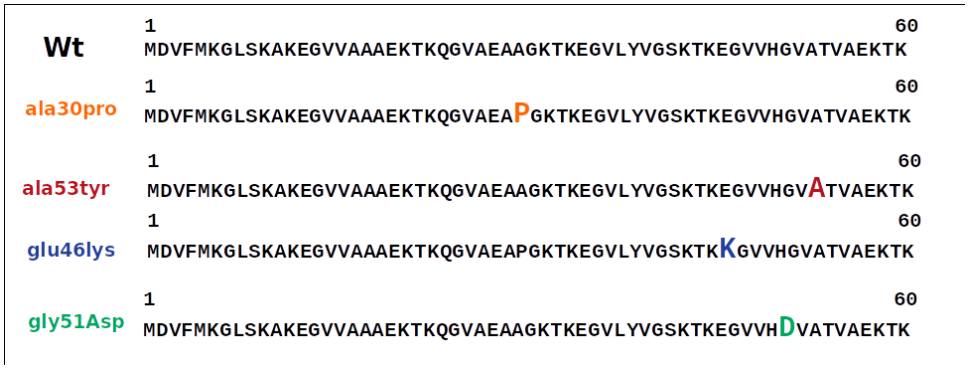
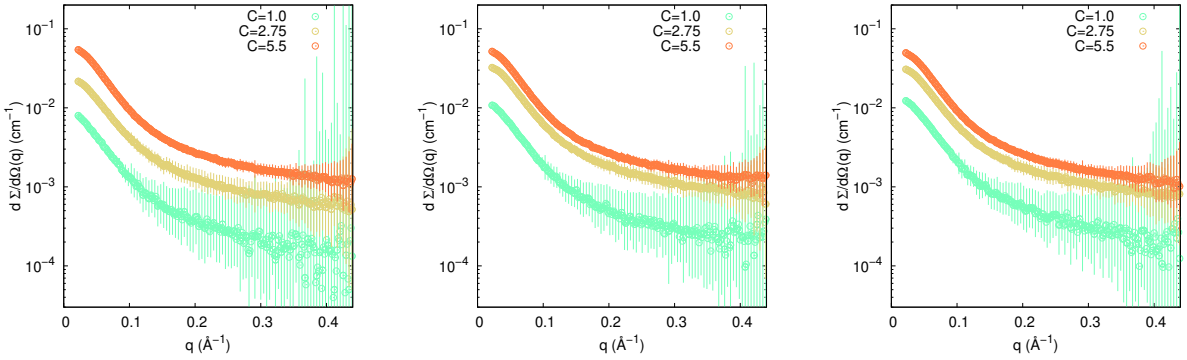


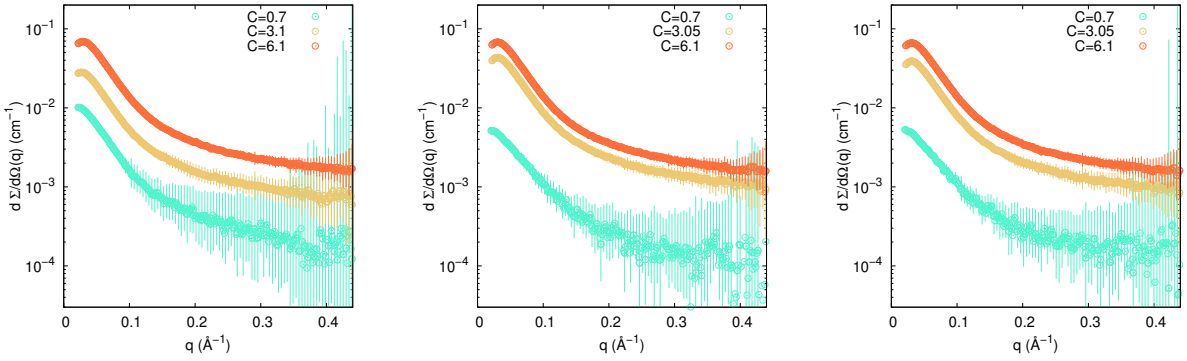
FIGURE 3.3: Point mutation in α-synuclein primary sequence in comparison with WT.

For each sample a small amount of $\sim$ 30 μL has been inserted in the capillary. It was also measured the SAXS pattern of the buffer (PBS pH 7.4) to be subtracted from the experimental curve. Figs. 3.4-3.8 shows some of the various curves recorded in different experimental sessions.



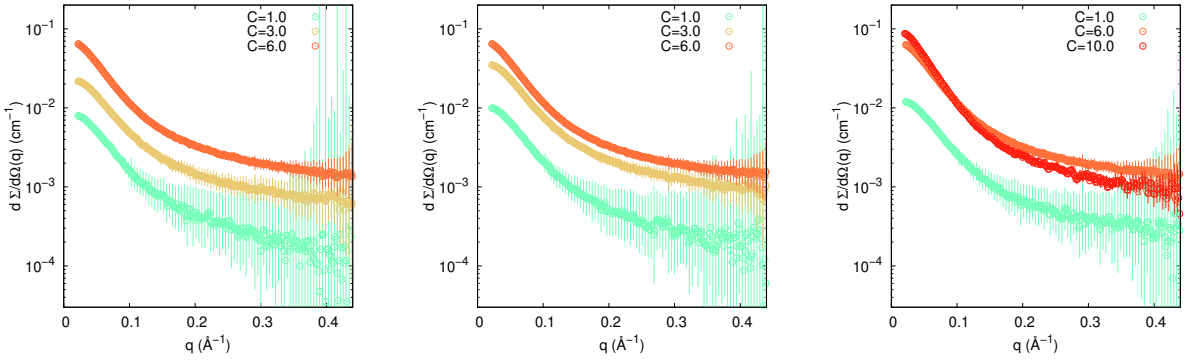
(a) $25°C$ (b) $37°C$ (c) $45°C$

FIGURE 3.4: SAXS curve of WT α-synuclein



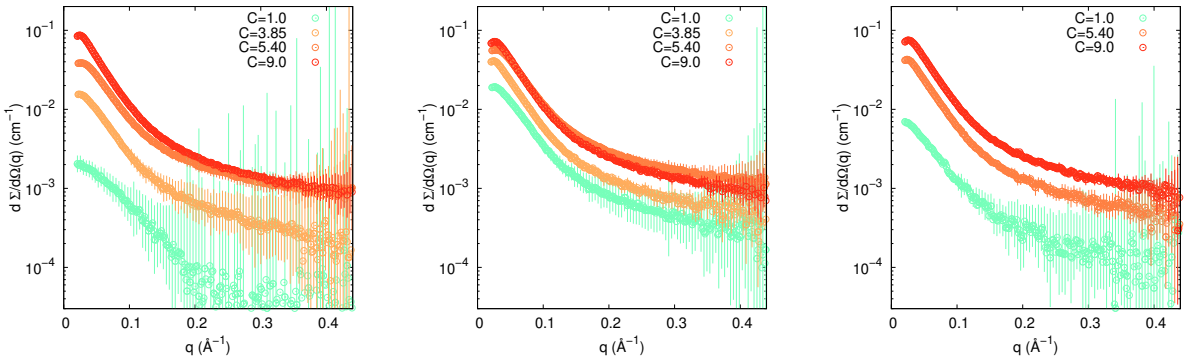
(a) $25°C$ (b) $37°C$ (c) $45°C$

FIGURE 3.5: SAXS curve of G51D α-synuclein mutant



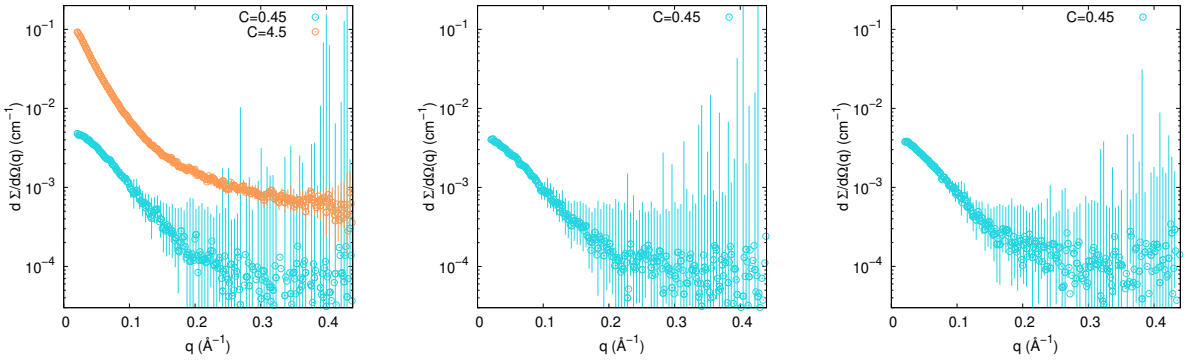
(a) $25°C$ (b) $37°C$ (c) $45°C$

FIGURE 3.6: SAXS curve of E46K α-synuclein mutant



(a) $25°C$ (b) $37°C$ (c) $45°C$

FIGURE 3.7: SAXS curve of A53T α-synuclein mutant



(a) $25°C$ (b) $37°C$ (c) $45°C$

FIGURE 3.8: SAXS curve of A30P α-synuclein mutant

Chapter 4

Method

4.1 VBW methods: from SAXS to propensities

The overall goal of this work is the description of the conformational order/disorder of IDPs in terms of folding propensities of each aminoacid. The idea is to develop a method able to obtain information about IDPs structural features from a combination of computational results (e.g. a conformational ensemble) and SAS experimental data, as introduced in Chapter 3. The focus of the method is the definition of *propensity*, that will be defined in the next paragraphs. Firstly, it is necessary to define some fundamental aspects:

- definition of the conformational space;
- availability of a conformational ensemble;
- availability of SAXS or SANS data to analyze (shown in Chapter 3).

4.2 Conformational space

The first problem addressed in this methodological study has been to find a criterion for dividing the space of the angles φ and ψ of Ramachandran [36] in regions that represent the most significant conformational structures. For example, according to Ozenne et al. [33] Ramachandran's plot is divided into 4 regions defined as α-left, α-right, β-proline and β -sheet, according to these relationships:

- αL $(\varphi > O^{\circ})$;
- αR $(\varphi < O^{\circ}, -120^{\circ} < \psi < 150^{\circ})$;
- βP $(-100^{\circ} < \varphi < O^{\circ}, \psi > 50^{\circ}, \psi < 120^{\circ})$
- βS $(-180^{\circ} < \varphi < -10O^{\circ}, \psi > 50^{\circ}, \psi < 120^{\circ})$

This choice seemed unrepresentative to us. Therefore the subdivision of the Ramachandran plan proposed in this thesis work was based on a probability distribution function of the φ and ψ angles obtained from a representative database of 500 structures present within the Protein Data Bank (*http://kinemage.biochem.duke.edu/databases/top500.php*). This database was created starting from 1000 protein structures and through the application of different structural analysis techniques the probability distribution $p(\varphi, \psi)$ of the mediated conformations on all the residues of all proteins was generated. Using this probability distribution function (Fig. 4.1) the Ramachandran plot has been divided into 8 regions, obtained by cutting the $p(\varphi, \psi)$ surface in the probability values of 0.0005 and 0.02, which define respectively energetically allowed and energetically favored regions [37].

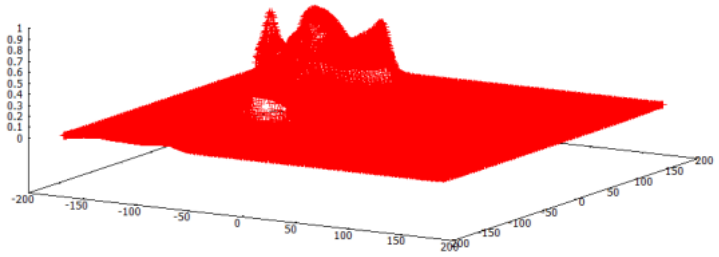


FIGURE 4.1: Probability distribution function from databases top500

The two-dimensional representation of the plot (Fig. 4.2 4.3) was obtained with the use of the Gnuplot software (*http://www.gnuplot.info/*).

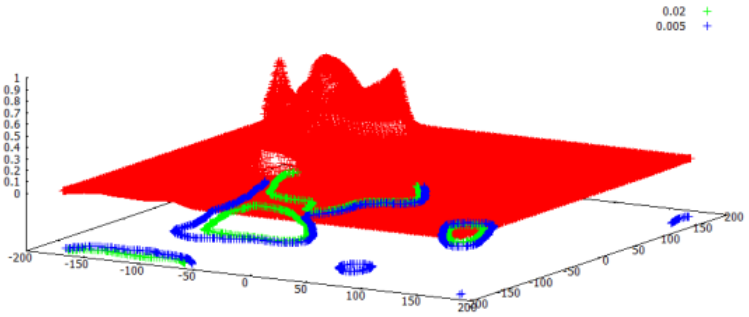


FIGURE 4.2: Visualization of the two-dimensional regions obtained from the distribution surface

The eight regions in which we decided to divide the plot are the 4 canonical regions in which there are the structures β, α-right, α-left and *glicine zone*, divided into energetically favorable, and energetically allowed, precisely as a function of the probability values with which the distribution was cut. Thus we have defined the following areas:

1. β favorable;
2. β allowed;
3. αR favorable;
4. αR allowed;

5. αL favorable;

6. αL allowed;

7. *glicine zone*;

8. *unstructured*.

These regions are clearly shown in Fig. 4.3

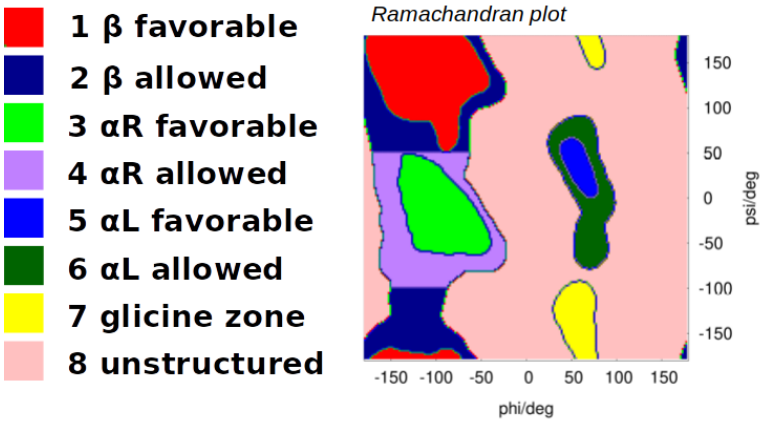


FIGURE 4.3: Visualization of Ramachandran's plot for the VBWSAS method

This was the subdivision designed to describe the conformational space in which the concept of propensity will be developed, which will be explained in the following sections.

4.3 Conformational ensembles

As explained in the Section 2.6, SAS is an ideal tool to integrate NMR and Crystallographic information to build complete structural and dynamic models of disordered biomolecular complexes [16]. This innovative method needs the use of conformational ensembles derived from computational results, in particular from Molecular Dynamics

simulations, in order to obtain structural features about the size and the overall shape of the IDPs. For this thesis work a conformational ensemble of 189 structure of α-synuclein has been chosen. These structures have been obtained by Gurry et al. [17] and are available in the Supporting Information of the paper. The authors, according to other studies, have suggested that α-syn can form soluble multimers in vivo that have significant secondary structure content. α-syn can indeed form β-strand-rich oligomers that are neurotoxic, and recent observations argue for the existence of soluble helical tetrameric structures that do not form toxic aggregates [17]. The Gurry's ensemble has been generated using data from NMR chemical shifts and residual dipolar couplings as guide. The dominant state of this ensemble is a disordered monomer, complemented by a small fraction of helical trimers and tetramers as well as by trimetric and tetrameric that are rich in β-strand content as shown in Fig. 4.4.

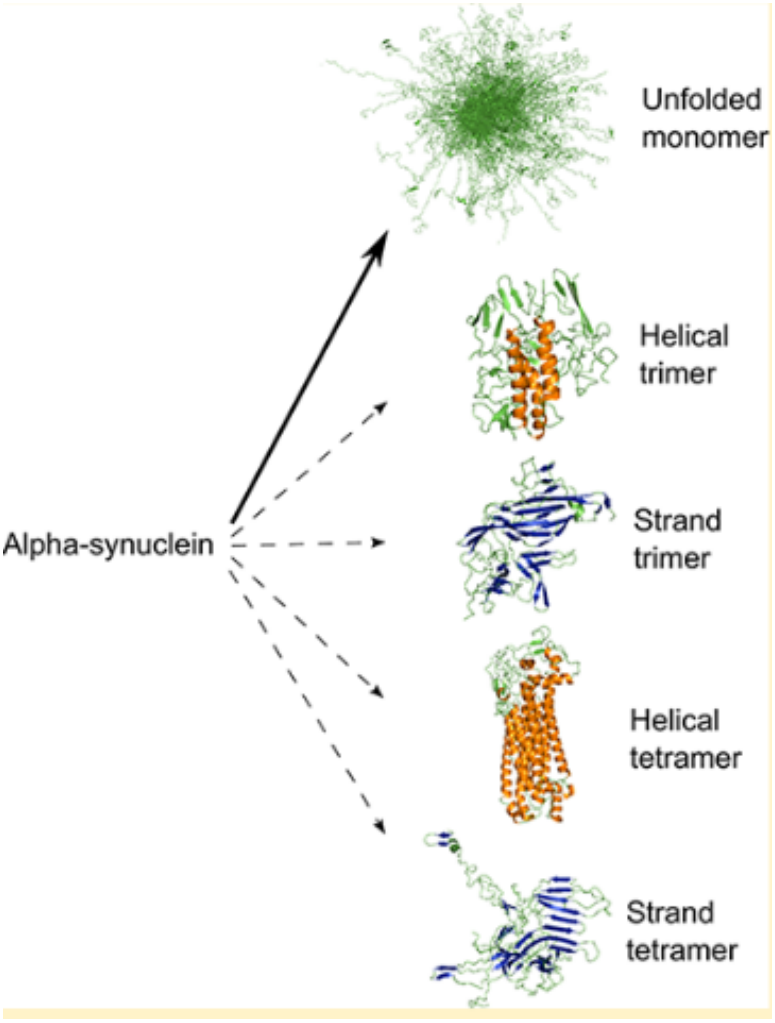


FIGURE 4.4: Gurry's ensemble [17]

The ensemble's composition is shown in Fig. 4.5.

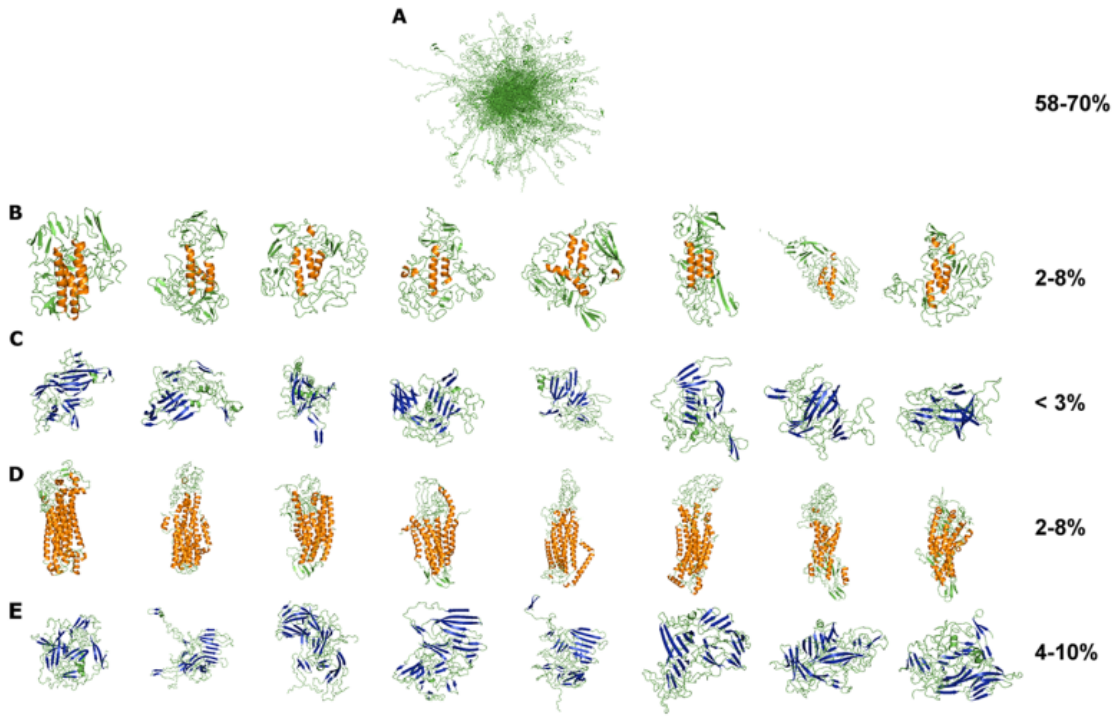


FIGURE 4.5: Gurry's ensemble composition: (A) Monomers; (B) Helical-rich trimers, (C) Strand-rich trimers, (D) Helical-rich tetramers, and (E) Strand-rich tetramers [17].

The multimer percentage composition of the ensemble results as:

- monomeric species, $(64 \pm 6)\%$;
- trimeric species, $(8 \pm 4)\%$;
- tetrameric species, $(28 \pm 6)\%$.

These are the main characteristics of the ensemble used in the SAS data analysis.

4.4 Propensity

Let us consider to have an ensemble of N conformations of the IDP under investigation, which is constituted by polypeptide chains of $N_{\rm aa}$ residues (aminoacids). We also assume that this ensemble contains all the conformational states that the IDP can adopt in any condition experimentally observed. As a consequence, at a given chemical-physical condition (e.g. concentration, temperature, pH etc.), the IDP will be distributed in the N conformations according to a set of relative weights w_i, with the normalization condition $\sum_{i=1}^{N} w_i = 1$. We define $\mathbf{w}$ the set of all the weights. Due to the intrinsic conformational disorder, the values of the set $\mathbf{w}$ could be distributed over a probability density function $P(\mathbf{w})$. Accordingly, the average value of w_i and of the product $w_{i_1} w_{i_2}$ are

$$< w_i > \;=\; \int w_i \, P(\mathbf{w}) \, d\mathbf{w} \tag{4.1}$$

$$< w_{i_1} w_{i_2} > \;=\; \int w_{i_1} w_{i_2} \, P(\mathbf{w}) \, d\mathbf{w} \tag{4.2}$$

As already discussed, the Ramachandran ψ, φ plane of the secondary structure of the protein can be divided in $N_{\rm reg} = 8$ finite regions. Hence, for each i-conformation, it possible to find the index of the region in which the Ramachandran angles of the a-residue (ranging from $a = 2$ to $a = N_{\rm aa} - 1$) falls. We also assume that each conformer could exist in a given oligomeric state and, accordingly, we define m_i the monomer aggregation number of the i-conformer. By referring to the entire set of conformations, the propensity

that the a-residue is in the r-region is defined according to

$$p_a(r) = \sum_{i=1}^{N} w_i \frac{1}{m_i} \sum_{g=1}^{m_i} \delta_{r,r_{iga}} \tag{4.3}$$

where r_{iga} is the index of the conformation region of the a-residue of the g-chain of the i-conformer, $\delta_{i,j}$ being Kronecker's delta function. The average propensity and its variance will be defined as

$$\mu_{p_a(r)} = \sum_{i=1}^{N} < w_i > \frac{1}{m_i} \sum_{g=1}^{m_i} \delta_{r,r_{iga}} \tag{4.4}$$

$$\sigma^2_{p_a(r)} = \sum_{i=1}^{N} (< w_i >^2 - < w_i^2 >) \frac{1}{m_i} \sum_{g_1=1}^{m_i} \delta_{r,r_{ig_1a}} \frac{1}{m_i} \sum_{g_2=1}^{m_i} \delta_{r,r_{ig_2a}}$$

$$+2 \sum_{i_1>i_2=1}^{N} (< w_{i_1} >< w_{i_2} > - < w_{i_1} w_{i_2} >) \frac{1}{m_{i_1}} \sum_{g_1=1}^{m_{i_1}} \delta_{r,r_{i_1g_1a}} \frac{1}{m_{i_2}} \sum_{g_2=1}^{m_{i_2}} \delta_{r,r_{i_2g_2a}} \tag{4.5}$$

4.4.1 Multidimensional Propensities

The concept of propensity can be extended to describe the probability to simultaneously have a set of K residues a_1, a_2,...,a_K in defined regions r_1, r_2,...,r_K,

$$p_{a_1,a_2,...,a_K}(r_1, r_2, \ldots, r_K) = \sum_{i=1}^{N} w_i \frac{1}{m_i} \prod_{k=1}^{K} \sum_{g=1}^{m_i} \delta_{r_k,r_{iga_k}} \tag{4.6}$$

In a more compact description, we can assign an index j to a given multidimensional propensity that can be seen as a function of the set $\mathbf{w}$, according to

$$
\begin{aligned}
p_j(\mathbf{w}) &= \sum_{i=1}^{N} w_i \, a_{ij} \\
a_{ij} &= \frac{1}{m_i} \prod_{k=1}^{K_j} \sum_{g=1}^{m_i} \delta_{r_k, r_{iga_k}} && (4.7) \\
\mu_{p_j} &= \sum_{i=1}^{N} < w_i > \; a_{ij} && (4.8) \\
\sigma^2_{p_j} &= \sum_{i=1}^{N} (< w_i^2 > - < w_i >^2) a_{ij}^2 && (4.9) \\
&\quad + 2 \sum_{i_1 > i_2}^{N} < w_{i_1} w_{i_2} > a_{i_1 j} a_{i_2 j} && (4.10)
\end{aligned}
$$

4.4.2 Propensities and Variational Bayesian Weighting

According to the fundamental theorem of Bayesian statistics, the probability $P(\mathbf{w})$ is written as

$$
\begin{aligned}
P(\mathbf{w}) &= \frac{1}{Z} P_{in}(\mathbf{w}) P_{ex}(\mathbf{w}) && (4.11) \\
Z &= \int d\mathbf{w} P(\mathbf{w}) && (4.12)
\end{aligned}
$$

where $P_{in}(\mathbf{w})$ is the so-called *intrinsic* probability distribution, $P_{ex}(\mathbf{w})$ is the *external* distribution, coming from some experimental observation and Z is the normalization factor. In the Variational Bayesian method [38], the intrinsic probability is written according to

a Dirichlet distribution of Jeffreys prior type [38],

$$P_{in}(\mathbf{w}) = \frac{\Gamma(\frac{1}{2}N)}{[\Gamma(\frac{1}{2})]^N} \prod_{i=1}^{N} w_i^{-\frac{1}{2}} \tag{4.13}$$

The external probability depends on the set of experimental information available on the system. Let us assume to know a given set of N_p *propensities* and their standard deviations. We have,

$$P_{ex}(\mathbf{w}) = \prod_{j=1}^{N_p} \frac{1}{\sigma_{p_j}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\mu_{p_j} - p_j(\mathbf{w})}{\sigma_{p_j}}\right]^2\right\} \tag{4.14}$$

Moreover, also the combined probability $P(\mathbf{w})$ is approximated by the a general Dirichlet distribution,

$$P(\mathbf{w}) \approx D(\mathbf{w}, \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{N}\Gamma(\alpha_i)} \prod_{i=1}^{N} w_i^{\alpha_i - 1} \tag{4.15}$$

which is fully defined by the set of real positive parameters $\boldsymbol{\alpha}$, with $\alpha_0 = \sum_{i=1}^{N} \alpha_i$. It is worth to recall some important properties of the Dirichlet distribution, which is a the

multivariate generalization of the beta distribution,

$$
\begin{aligned}
< w_i > &= \frac{\alpha_i}{\alpha_0} && (4.16)\\
< w_i^n > &= \frac{\prod_{k=0}^{n-1}(\alpha_i + k)}{\prod_{k=0}^{n-1}(\alpha_0 + k)} && (4.17)\\
< w_i w_j > &= \begin{cases} \frac{\alpha_i \alpha_j}{\alpha_0(\alpha_0+1)} & i \neq j \\ \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} & i = j \end{cases} && (4.18)\\
\mathrm{cov}(w_i, w_j) &= < w_i w_j > - < w_i >< w_j >= \frac{\alpha_i \alpha_0 \delta_{i,j} - \alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} && (4.19)
\end{aligned}
$$

According to the Maximum Entropy principle, it has been shown that the set $\boldsymbol{\alpha}$ can by found by minimizing the functional

$$
\begin{aligned}
L(\boldsymbol{\alpha}) &= -\int d\mathbf{w} D(\mathbf{w}, \boldsymbol{\alpha}) \log \frac{P_{in}(\mathbf{w}) P_{ex}(\mathbf{w})}{D(\mathbf{w}, \boldsymbol{\alpha})} \\
&= -\int d\mathbf{w} D(\mathbf{w}, \boldsymbol{\alpha}) \log \frac{P_{in}(\mathbf{w})}{D(\mathbf{w}, \boldsymbol{\alpha})} \\
&\quad -\int d\mathbf{w} D(\mathbf{w}, \boldsymbol{\alpha}) \log P_{ex}(\mathbf{w}) && (4.20)
\end{aligned}
$$

The first term of Eq. 4.20 is found to be

$$
\begin{aligned}
&-\int d\mathbf{w} D(\mathbf{w}, \boldsymbol{\alpha}) \log \frac{P_{in}(\mathbf{w})}{D(\mathbf{w}, \boldsymbol{\alpha})} = \log \frac{\Gamma(\alpha_0)}{\Gamma(\frac{1}{2}N)} + \sum_{i=1}^{N} \log \frac{\Gamma(\frac{1}{2})}{\Gamma(\alpha_i)} \\
&+ \sum_{i=1}^{N} \left(\alpha_i - \frac{1}{2} \right) [\psi(\alpha_i) - \psi(\alpha_0)] && (4.21)
\end{aligned}
$$

being $\psi(x)$ the digamma function, $\psi(x) = \Gamma'(x)/\Gamma(x)$, $\Gamma(x)$ being the gamma function.

The second part of Eq. 4.20 can be expanded:

$$
\begin{aligned}
&-\int d\mathbf{w} D(\mathbf{w},\boldsymbol{\alpha}) \log P_{ex}(\mathbf{w}) = \sum_{j=1}^{N_p} \frac{1}{2\sigma_j^2} \int d\mathbf{w} D(\mathbf{w},\boldsymbol{\alpha}) \{\mu_{p_j} - p_j(\mathbf{w})\}^2 + C \\
&= \sum_{j=1}^{N_p} \frac{1}{2\sigma_j^2} \int d\mathbf{w} D(\mathbf{w},\boldsymbol{\alpha}) \left\{ \mu_{p_j} - \sum_{i=1}^{N} w_i a_{ij} \right\}^2 + C \\
&= \sum_{j=1}^{N_p} \frac{[\mu_{p_j} - \mu_{p_j}^{(\mathrm{fit})}]^2}{2\sigma_j^2} + \sum_{i_1 i_2 = 1}^{N} [< w_{i_1} w_{i_2} > - < w_{i_1} >< w_{i_2} >] \sum_{j=1}^{N_p} \frac{a_{i_1 j} a_{i_2 j}}{2\sigma_j^2} + C \quad (4.22)
\end{aligned}
$$

where C is a constant that depends on experimental standard deviations,

$$
C = -\sum_{j=1}^{N_p} \log[\sigma_j \sqrt{2\pi}]. \quad (4.23)
$$

To simplify the notation, it is useful to introduce several parameters. First, the standard chi square of the data is defined,

$$
\chi^2 = \frac{1}{N_p} \sum_{j=1}^{N_p} \frac{[\mu_{p_j} - \mu_{p_j}^{(\mathrm{fit})}]^2}{\sigma_j^2} \quad (4.24)
$$

The second term includes the covariance matrix of $\mathbf{w}$

$$
\begin{aligned}
A = &\sum_{i_1=1}^{N} \left[\frac{\alpha_{i_1}(\alpha_{i_1}+1)}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)^2 \right] A_{i_1 i_1} \\
&+2 \sum_{i_1 < i_2}^{N} \left[\frac{\alpha_{i_1}\alpha_{i_2}}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)\left(\frac{\alpha_{i_2}}{\alpha_0}\right) \right] A_{i_1 i_2} \quad (4.25)
\end{aligned}
$$

where

$$A_{i_1 i_2} = \frac{1}{N_p} \sum_{j=1}^{N_p} \frac{a_{i_1 j} a_{i_2 j}}{2\sigma_j^2} \tag{4.26}$$

According to Eq. 4.5, the average propensity and its standard deviation become

$$\mu_{p_j}^{(\mathrm{fit})} = \sum_{i=1}^{N} \frac{\alpha_i}{\alpha_0} a_{ij} \tag{4.27}$$

$$\begin{aligned} \sigma_{p_j}^{2(\mathrm{fit})} = & \sum_{i_1=1}^{N} \left[\frac{\alpha_{i_1}(\alpha_{i_1}+1)}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)^2 \right] a_{i_1 j}^2 \\ & + 2 \sum_{i_1<i_2}^{N} \left[\frac{\alpha_{i_1}\alpha_{i_2}}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)\left(\frac{\alpha_{i_2}}{\alpha_0}\right) \right] a_{i_1 j} a_{i_2 j} \end{aligned} \tag{4.28}$$

Combining the last equation with Eq. 4.21, the final expression of $L(\boldsymbol{\alpha})$ results

$$\begin{aligned} L(\boldsymbol{\alpha}) = & \log \frac{\Gamma(\alpha_0)}{\Gamma(\frac{1}{2}N)} + \sum_{i=1}^{N} \log \frac{\Gamma(\frac{1}{2})}{\Gamma(\alpha_i)} \\ & + \sum_{i=1}^{N} \left(\alpha_i - \frac{1}{2}\right) [\psi(\alpha_i) - \psi(\alpha_0)] + \frac{N_p}{2}(\chi^2 + A) + C \end{aligned} \tag{4.29}$$

In summary, when a set of propensity and their standard deviations is known, by minimizing $L(\boldsymbol{\alpha})$ an optimum set of Dirichlet parameter $\boldsymbol{\alpha}$ can be found and then average and standard deviation values of any conformational weight w_i are determined.

4.4.3 SAS and Variational Bayesian Weighting

The Variational Bayesian Weighting formalism can in principle be applied to any experimental observation obtained over a system of IDPs. In the case of small-angle X-ray as well as neutron scattering data, the external probability is represented by the following equation

$$P_{ex}(\mathbf{w}) = \prod_{k=1}^{N_q} \frac{1}{\sigma(q_k)\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{\frac{d\Sigma}{d\Omega}^{\exp}(q_k) - \frac{d\Sigma}{d\Omega}(q_k, \mathbf{w})}{\sigma(q_k)} \right]^2 \right\} \tag{4.30}$$

where, $\frac{d\Sigma}{d\Omega}^{\exp}(q_k)$ and $\sigma(q_k)$ represents the experimental macroscopic differential X-ray or neutron scattering cross section and its standard deviation, respectively, measured in the k-th of N_q values of the scattering vector modulus q_k.

In the most general case, let us assume that the IDP molecules, with conformations described by a selected ensemble, can have a relatively high concentration so that in the SAS curve also the effect of long range protein-protein interaction can be depicted. In these conditions, we first introduce the nominal number density of monomers, $n_\circ$, and the number density n_i of any conformer according to

$$n_\circ = \frac{cN_A}{M_1} \tag{4.31}$$

$$n_i = n_\circ \frac{w_i}{m_i} \tag{4.32}$$

where M_1 is the monomer molecular weight, N_A Avogadro's number, c the nominal w/v protein concentration and, as already said, m_i is the aggregation number of the i

multimer. Considering a unique average protein-protein structure factor $S(q)$ that takes into account interaction effect among any conformation, according to the SAS formalism, the macroscopic differential scattering cross section can be seen as a function of the set $\mathbf{w}$,

$$\frac{d\Sigma}{d\Omega}(q,\mathbf{w}) = \sum_{i=1}^{N} w_i a_i(q) + B \tag{4.33}$$

$$a_i(q) = \frac{n_\circ}{m_i} F_i^{(2)}(q)\{1+\beta_{\text{ell}}(q)[S(q)-1]\}, \tag{4.34}$$

so that the fitting curve is written in term of the average populations,

$$\frac{d\Sigma}{d\Omega}^{\text{fit}}(q) = \frac{d\Sigma}{d\Omega}(q,<\mathbf{w}>) \tag{4.35}$$

In equation 4.34, $F_i^{(2)}(q)$ is the squared form factor of the i-conformer, that can be calculated on the basis of atomic coordinates (e.g from a PDB file) for both X-rays or neutron scattering by means of methods such as SASMOL (see Fig. 2.9).

$\beta_{\text{ell}}(q)$ plays the role of the ratio $|<F_{\text{eff}}(\mathbf{q})>|^2/<|F_{\text{eff}}(\mathbf{q})|^2>$ between the square of the effective orientational average form factor and the orientational average of the effective squared form factor. As discussed by Pedersen et al. [39], this function, which typically deviates from 1 for anisometric shapes, can be approximated in an acceptable way by assuming that the effective particle has a simple geometrical shape. In our case, we have considered the shape of a biaxial ellipsoid.

$S(q)$ is calculated as the perturbation of the hard sphere structure factor $S_0(q)$ obtained with the well-known Percus-Yevick (PY) approximation in the framework of the random

phase approximation (RPA). The perturbation is due to the presence of two Yukawian terms, the first representing the screened Coulumbian repulsion potential and the other an attractive potential [40]. The relevant parameters of this approximation are the effective values of the number density, n, the particle radius, R, the net charge, Z, the inverse Debye screening length, κ_D (which depends on the ionic strength I_S of the protein solution), the attractive potential at contact, J, and the range of the attractive interaction, d. Considering the average aggregation number

$$\mu_m = \sum_{i=1}^{N} m_i < w_i >, \tag{4.36}$$

the parameters of the effective protein-protein structure factor can be written as follows

$$n = n_\circ \sum_{i=1}^{N} < w_i > / m_i \tag{4.37}$$

$$Z = Z_1 \mu_m \tag{4.38}$$

$$J = J_1 \mu_m^{2/3} \tag{4.39}$$

$$R = R_1 \mu_m^{1/3} \tag{4.40}$$

where Z_1, J_1 and R_1 are the monomer charge, attractive depth and average radius. Notice that we have supposed that Z is simply proportional to μ_m. On the other hand, J is supposed to scale as the surface of the protein, here simply defined as the one of the sphere defined by the radius R. This latter clearly scales as the cubic root of the volume, which is directly proportional to μ_m.

According to this view, the volume of the biaxial ellipsoid that is used to determine

$\beta_{\rm ell}(q)$ is $(4/3)\pi R_1^3\mu_m$: hence the only parameter that should be optimized is the ellipsoid anisometry ν, i.e. the ratio between the semi-axis a and b, b representing the two equal semi-axes.

In equation 4.33, B is flat background which takes into account incoherent scattering effects, particularly relevant in SANS experiments.

The VBW formalism follows a way similar to the one used for the application of propensities data. After some algebra, it can be shown that the potential to be minimized over the space of the Dirichlet set $\boldsymbol{\alpha}$ is

$$\begin{aligned} L(\boldsymbol{\alpha}) &= \log\frac{\Gamma(\alpha_0)}{\Gamma(\frac{1}{2}N)} + \sum_{i=1}^{N}\log\frac{\Gamma(\frac{1}{2})}{\Gamma(\alpha_i)} \\ &\quad + \sum_{i=1}^{N}\left(\alpha_i - \frac{1}{2}\right)[\psi(\alpha_i) - \psi(\alpha_0)] + \frac{N_q}{2}(\chi^2 + A) + C \end{aligned} \tag{4.41}$$

where

$$\chi^2 = \frac{1}{N_q}\sum_{k=1}^{N_q}\frac{[\frac{d\Sigma}{d\Omega}^{\rm exp}(q_k) - \frac{d\Sigma}{d\Omega}^{\rm fit}(q_k)]^2}{\sigma^2(q_k)} \tag{4.42}$$

$$\begin{aligned} A &= \sum_{i_1=1}^{N}\left[\frac{\alpha_{i_1}(\alpha_{i_1}+1)}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)^2\right]A_{i_1 i_1} \\ &\quad + 2\sum_{i_1<i_2}^{N}\left[\frac{\alpha_{i_1}\alpha_{i_2}}{\alpha_0(\alpha_0+1)} - \left(\frac{\alpha_{i_1}}{\alpha_0}\right)\left(\frac{\alpha_{i_2}}{\alpha_0}\right)\right]A_{i_1 i_2} \end{aligned} \tag{4.43}$$

$$A_{i_1 i_2} = \frac{1}{N_q}\sum_{k=1}^{N_q}\frac{a_{i_1}(q_k)a_{i_2}(q_k)}{\sigma^2(q_k)} \tag{4.44}$$

$$C = -\sum_{k=1}^{N}{}_q\log[\sigma^2(q_k)\sqrt{2\pi}] \tag{4.45}$$

The overall method here introduced has been named VBWSAS, version 1.0. To note, the application of this method to the analysis of a SAS curve includes the optimization not only of the $N+1$ Dirichlet parameters of the set $\boldsymbol{\alpha}$ but also all the parameters related to the form factor, (in particular the relative mass density $d_{\mathrm{W,i}}$ of the hydration water of the i-th conformer), and the ones of the effective structure factor (the monomer radius, charge, attractive potential and range, as well as the ionic strength and the ellipsoid anisometry).

4.4.4 Multiple Equilibria of IDPs

From a thermodynamic point of view, in an ideal solution, the chemical potential of any monomeric chain of the intrinsically disordered protein in the i-conformation forming a multimer with m_i aggregation state is defined as

$$\mu_i = \mu_{\circ i} + \frac{RT}{m_i} \log C_i, \tag{4.46}$$

where R is the universal gas constant, T the absolute temperature and C_i the molar concentration of the i-multimer, corresponding to $C_i = (c/M_1)(< w_i > /m_i)$. At the equilibrium, the chemical potentials of any monomer are equal. Hence, by referring the the first multimer, we have

$$\log C_i \;\; = \;\; -\frac{m_i \Delta G_{i1}}{RT} + \frac{m_i}{m_1} \log C_1, \tag{4.47}$$

where $\Delta G_{i1} = \mu_{\circ i} - \mu_{\circ 1}$ is the standard Gibbs free energy change corresponding to the transformation of a solution 1 M of monomers in the 1-conformation and having aggregation number m_1 into monomers in the i-conformation with aggregation number m_i. The last equation allows to derive the number fraction of monomers in the i-conformer as as function of the one of the first conformer,

$$< w_i > = m_i (c/M_1)^{\frac{m_i}{m_1}-1} (< w_1 > /m_1)^{\frac{m_i}{m_1}} e^{-m_i \Delta G_{i1}/(RT)} \tag{4.48}$$

By combining with the normalization conditions $\sum_{i=1}^{N} < w_i >= 1$, a polynomial equation of degree $\gamma = \max\{m_i\}$ of the unique variable $x =< w_1 >^{1/m_1}$ is obtainable,

$$\sum_{j=1}^{\gamma} A_j x^j - 1 = 0 \tag{4.49}$$

$$A_j = j(c/M_1)^{\frac{j}{m_1}-1} (1/m_1)^{\frac{j}{m_1}} \sum_{i=1}^{N} \delta_{j,m_i} e^{-j\Delta G_{i1}/(RT)}, \tag{4.50}$$

Where δ_{ij} if the Kronecker delta function. According to Abel–Ruffini theorem, analytic solutions are available only up to $\gamma = 4$, i.e. up to the formation of tetramers, which are the multimers with the maximum aggregation number in the Gurry's ensemble.

Classical thermodynamic allows also to describe the Gibbs free energy change as a function of T in terms of standard enthalpy, standard entropy and heat capacity at constant pressure variations, all referred to the first conformer,

$$\Delta G_{i1} = \Delta H^0_{i1} - T\Delta S^0_{i1} + \Delta C_{p_{i1}} \left[(T - T_0) - T \log \frac{T}{T_0} \right] \tag{4.51}$$

where T_0 is a standard temperature, typically 298 K. To note, the ratio $\frac{\Delta G_{i1}}{RT}$, seen in Eq. 4.50, can be written in terms of three dimensionless variations of enthalpy, $\frac{\Delta H^0_{i1}}{RT_0}$, entropy $\frac{\Delta S^0_{i1}}{R}$ and heat capacity $\frac{\Delta C_{p_{i1}}}{R}$

$$\frac{\Delta G_{i1}}{RT} = \frac{T_0}{T}\frac{\Delta H^0_{i1}}{RT_0} - \frac{\Delta S^0_{i1}}{R} + \frac{\Delta C_{p_{i1}}}{R}\left[1 - \frac{T_0}{T} - \log\frac{T}{T_0}\right] \tag{4.52}$$

This thermodynamic approach can be integrated in the VBWSAS in order to analyze not only one SAS curve but several curves, recorded on the same IDP as a function, for example, of whole protein concentration c and temperature. Indeed, with optimum values of the three parameters $\frac{\Delta H^0_{i1}}{RT_0}$, $\frac{\Delta S^0_{i1}}{R}$ and $\frac{\Delta C_{p_{i1}}}{R}$ the values of $<w_i>$, according to Eq. 4.48, can be found. Given a value of α_0 for each SAS curve, according to Eq. 4.16, from $<w_i>$ we can derive the values of α_i relative to a SAS curve. Hence a unique minimization of all the functionals $L(\boldsymbol{\alpha})$ for each SAS curve can be indirectly carried out not in the space of $\boldsymbol{\alpha}$ but in the space of the thermodynamic parameters of any conformer and of α_0 for any SAS curve. We have named this approach VBWSAS 2.0.

4.4.5 VBW on different classes of multimers

The application of VBWSAS 2.0 to the set of SAS curves of α-syn did not provide good fitting results. Hence, a further modification of the Variational Bayesian Weighting has been developed. The idea is that the ensemble containing N conformers can be subdivided in M classes of multimers having the same aggregation number. Let us assume that in the m^{th} class there are N_m conformers, with $\sum_{m=1}^{M} N_m = N$. We introduce the set $\mathbf{W}_m$ that

contains *the conformers' probabilities* of the class, with the straightforward normalization condition

$$\sum_{j=1}^{N_m} W_{m\,j} = 1 \tag{4.53}$$

Hence, if the protein monomer is in the m^{th} class, $W_{m\,j}$ represents the probability that it is folded according to the j^{th} conformer of that class. The probability to find the protein in the m^{th} class, i.e. to find the protein with the aggregation number m independently on the conformer within the class, is indicated by the symbol ω_m, with the normalization condition

$$\sum_{m=1}^{M} \omega_m = 1 \tag{4.54}$$

With these definitions, the probability w_i of the i^{th} conformer among all the N conformers of the ensemble is

$$w_i = \omega_{m_i} W_{m_i\,j_i} \tag{4.55}$$

where m_i is the class to which the i-conformer belongs and j_i indicates to which of the N_{m_i} conformers of that class the i-conformer corresponds.

By generalizing the Variational Bayesian method, we introduce the probability distribution $P(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ to find out a set of probability vectors $\mathbf{W}_1, \ldots, \mathbf{W}_M$ as

$$P(\mathbf{W}_1, \ldots, \mathbf{W}_M) = \frac{1}{Z} P_{in}(\mathbf{W}_1, \ldots, \mathbf{W}_M) P_{ex}(\mathbf{W}_1, \ldots, \mathbf{W}_M) \tag{4.56}$$

$$Z = \int d\mathbf{W}_1, \ldots, d\mathbf{W}_M P(\mathbf{W}_1, \ldots, \mathbf{W}_M) \tag{4.57}$$

where $P_{in}(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ is the so-called *intrinsic* probability distribution, $P_{ex}(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ is the *external* distribution, coming from some experimental observation and Z is the normalization factor.

Now, in this new approach of VBWSAS we make the assumption that the statistical conformational probability distribution within each class is independent so that an intrinsic distribution of the class, $P_{in\,m}(\mathbf{W}_m)$, can be introduced; moreover, according to the VBW method, we write this distribution by a Dirichlet function

$$P_{in\,m}(\mathbf{W}_m) = \frac{\Gamma(\frac{1}{2}N_m)}{[\Gamma(\frac{1}{2})]^{N_m}} \prod_{j=1}^{N_m} W_{m\,j}^{-\frac{1}{2}} \tag{4.58}$$

Hence the overall intrinsic distribution is the product of all the M class intrinsic distributions,

$$P_{in}(\mathbf{W}_1, \ldots, \mathbf{W}_M) = \prod_{m=1}^{M} P_{in\,m}(\mathbf{W}_m) \tag{4.59}$$

Moreover, we make the stronger assumption that *also* the whole probability distribution $P(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ can be seen as the product of M Dirichlet functions

$$P(\mathbf{W}_1, \ldots, \mathbf{W}_M) \approx \prod_{m=1}^{M} D_m(\mathbf{W}_m, \boldsymbol{\alpha}_m) = \prod_{m=1}^{M} \frac{\Gamma(\alpha_{m\,0})}{\prod_{j=1}^{N_m} \Gamma(\alpha_{m\,j})} \prod_{j=1}^{N_m} W_{m\,j}^{\alpha_{m\,j}-1} \quad (4.60)$$

with $\alpha_{m\,0} = \sum_{j=1}^{N_m} \alpha_{m\,j}$.

On the basis of the Dirichlet distribution properties, the average sequence $< \mathbf{W}_m >$ of conformers' probability of the class is associated to the sequence $\boldsymbol{\alpha}_m$ of Dirichlet parameters, with $< W_{m\,j} >= \alpha_{m\,j}/\alpha_{m\,0}$. It can be demonstrated that all the sequence of Dirichlet parameters $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M$ can be found by minimizing the functional

$$L(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M) = -\sum_{m=1}^{M} \int d\mathbf{W}_m D_m(\mathbf{W}_m, \boldsymbol{\alpha}_m) \log \frac{P_{in\,m}(\mathbf{W}_m)}{D_m(\mathbf{W}_m, \boldsymbol{\alpha}_m)}$$
$$- \int d\mathbf{W}_1 \ldots d\mathbf{W}_M \left(\prod_{m=1}^{M} D_m(\mathbf{W}_m, \boldsymbol{\alpha}_m) \right) \log P_{ex}(\mathbf{W}_1, \ldots, \mathbf{W}_M) \quad (4.61)$$

The external probability distribution depends on the experimental information provided by SAS according to

$$P_{ex}(\mathbf{W}_1, \ldots, \mathbf{W}_M) = \prod_{k=1}^{N_q} \frac{1}{\sigma(q_k)\sqrt{2\pi}}$$
$$\times \exp \left\{ -\frac{1}{2} \left[\frac{\frac{d\Sigma}{d\Omega}^{\exp}(q_k) - \frac{d\Sigma}{d\Omega}(q_k, \mathbf{W}_1, \ldots, \mathbf{W}_M)}{\sigma(q_k)} \right]^2 \right\} \quad (4.62)$$

where the calculated macroscopic scattering cross sections, which have to be close to the corresponding experimental values provided by SAS, is expressed as a function of $\boldsymbol{W}_m$

and ω_m according to

$$\frac{d\Sigma}{d\Omega}(q, \mathbf{W}_1, \ldots, \mathbf{W}_M) = \sum_{m=1}^{M} \omega_m \sum_{j=1}^{N_m} W_{m\,j} a_{m\,j}(q) + B \tag{4.63}$$

$$= \sum_{i=1}^{N} w_i a_i(q) + B \tag{4.64}$$

The Eq. 4.61 transforms to

$$\begin{aligned} L(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M) = & \sum_{m=1}^{M} \left\{ \log \frac{\Gamma(\alpha_{m\,0})}{\Gamma(\frac{1}{2}N_m)} + \sum_{j=1}^{N_m} \log \frac{\Gamma(\frac{1}{2})}{\Gamma(\alpha_{m\,j})} \right. \\ & + \sum_{j=1}^{N_m} \left(\alpha_{m\,j} - \frac{1}{2} \right) [\psi(\alpha_{m\,j}) - \psi(\alpha_{m\,0})] \\ & + \frac{N_q}{2} \left[\sum_{j=1}^{N_m} \left[\frac{\alpha_{m\,j}(\alpha_{m\,j}+1)}{\alpha_{m\,0}(\alpha_{m\,0}+1)} - \left(\frac{\alpha_{m\,j}}{\alpha_{m\,0}} \right)^2 \right] A_{m\,j\,j} \right. \\ & \left. \left. + 2 \sum_{j_1 > j_2 = 1}^{N_m} \left[\frac{\alpha_{m\,j_1}\alpha_{m\,j_2}}{\alpha_{m\,0}(\alpha_{m\,0}+1)} - \frac{\alpha_{m\,j_1}}{\alpha_{m\,0}} \frac{\alpha_{m\,j_2}}{\alpha_{m\,0}} \right] A_{m\,j_1\,j_2} \right] \right\} \\ & + \frac{N_q}{2}\chi^2 - \sum_{k=1}^{N_q} \log[\sigma(q_k)\sqrt{2\pi}] \end{aligned} \tag{4.65}$$

with the following definitions

$$A_{m\,j_1\,j_2} = \frac{1}{N_q} \sum_{k=1}^{N_q} \frac{a_{m\,j_1}(q_k) a_{m\,j_2}(q_k)}{\sigma^2(q_k)} \tag{4.66}$$

$$\chi^2 = \frac{1}{N_q} \sum_{k=1}^{N_q} \frac{[\frac{d\Sigma}{d\Omega}^{\exp}(q_k) - \frac{d\Sigma}{d\Omega}(q_k, <\mathbf{W}_1>, \ldots, <\mathbf{W}_M>)]^2}{\sigma^2(q_k)} \tag{4.67}$$

It is important to notice that the VBW method is now applied only within each class. Indeed, the overall set of SAS curves used to fit the data depend on the class probability

ω_m. On the other hand, according to the thermodynamic model, we are able to calculate all the values of $<w_i>$ (Eq. 4.48) and then we can derive the values of ω_m,

$$\omega_m = \sum_{i=1}^{N} \delta_{m_i,m} <w_i> \tag{4.68}$$

and the values of $<W_{m\,j}>$,

$$<W_{m\,j}> = \frac{<w_{i_{m\,j}}>}{\omega_m} \tag{4.69}$$

where $i_{m\,j}$ is the conformer among the ensemble of N conformers corresponding to the j^{th} conformer of the m^{th} class. On the other hand, the Dirichlet parameters $\boldsymbol{\alpha}_m$ can be expressed as a function of $<\mathbf{W}_m>$ and $\alpha_{m\,0}$ according to

$$\alpha_{m\,j} = \alpha_{m\,0} <W_{m\,j}> \tag{4.70}$$

In conclusion, the minimization of the functional $L(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M)$ can be performed by expressing the values of $\boldsymbol{\alpha}_m$ in terms of the free parameters of the thermodynamic model, which are the dimensionless standard enthalpy and standard entropy changes with respect to the first conformer of the ensemble ($\Delta H^0_{i1}/(RT_0)$ and $\Delta S^0_{i1}/R$, respectively) as well as the corresponding change of the dimensionless specific heat at constant pressure ($\Delta C_{p_{i1}}/R$), and in terms of the Dirchlet parameters $\alpha_{m\,0}$ of each class. Moreover the other free parameters that enter in the definition of $\frac{d\Sigma}{d\Omega}(q)$ are the ones related to the structure factor.

On these basis, if several SAS curves of the IDP of interest have been measured in different conditions of total weight concentration c and temperature T, according to the thermodynamic equilibria approach (Eq. 4.48), the values of $< w_i >$ can be calculated as a unique function of $\Delta H^0_{i1}/(RT_0)$, $\Delta S^0_{i1}/R$ and $(\Delta C_{p_{i1}}/R)$. Hence a unique best-fit of the entire set of SAS data can be realized by minimizing the sum over each SAS curve of the corresponding functional L (Eq. 4.65), where the same values of the thermodynamic parameters as well as the structure factor parameters are optimized of all the curves, whereas specific values of $\alpha_{m\,0}$ are optimized for each SAS curve.

We define this new formalism VBWSAS 3.0. As shown in the next chapter, with this approach we have been able to derive good fit results of SAS data.

As a final remark, we point out that original Fortran codes for the implementation of any version of VBWSAS have been written in the course of this thesis.

Chapter 5

Simulations

The aim of the work was to develop a computational tool that could calculate the propensity of amino acids of an IDPs starting from a SAS curve. Once the theoretical part of program's design ended, simulations was performed to verify the correct functioning. In the next paragraphs the validation of the process will be explained.

5.1 Method validation

In this section it will be explained how the VBWSAS 1.0 method was validated through the use of simulations. In the validation of the method, in order to reduce computational calculation time, an ensemble consisting of 100 conformations of β - amyloid was chosen, an IDP protein also involved in neurodegenerative diseases but composed of only 40 amino acids unlike the 140 of the α - synuclein. The Aβ ensemble chosen for the method's validation is composed as follow:

- 50 conformations generated by a home made software that randomly samples the function $f(\varphi, \psi)$ that as a single constraint had a control to avoid the overlap due to the steric encumbrance of the amino acids;

- 50 conformations taken from Molecular Dynamics simulation of the Aβ monomer after opportune balancing. The calculations were performed at the Department of Physics of the Tor Vergata University of Rome and kindly provided to us by Dr. Velia Minicozzi.

The method has been validated by testing its reliability in two particular cases that correspond to the two most important regions of the Ramachandran plot:

- β_{fav} region;

- αR_{fav} region.

5.1.1 β_{fav} region test

The first step of the work was to evaluate the average propensity ($\mu_{p_a(r)}$ of Eq. 4.5, called FLAT propensity) of the ensemble (Fig. 5.1) used as described in the section 4.4.

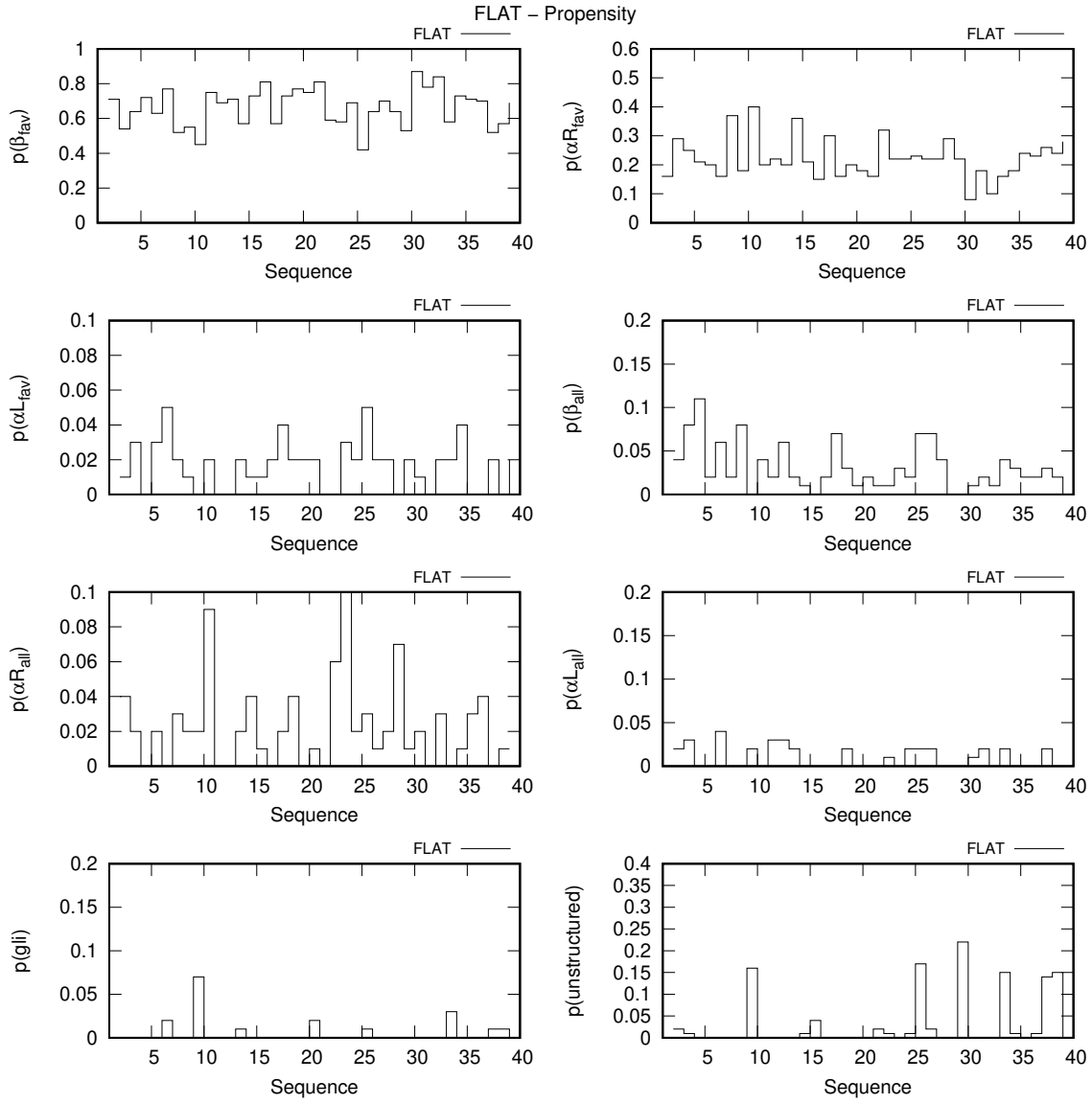


FIGURE 5.1: Averaged propensity of method's validation ensemble

The results show a high degree of disorder with a slightly higher β propensity than the α propensity. In order to test the correctness of the VBWSAS 1.0 method, the propensity of the amino acids 20-27 and 28-33 to be in the β_{fav} region was prefixed at values 0.8 and 1 with the aim of simulating a similar situation of the beginning of the aggregation process. This was done by calculating the weight factors, w_i, and their standard deviations, in order to reproduce a set of know average propensities $\mu_{p_a(r)}$ with their SDs, $\sigma^2_{p_a(r)}$, to simulate this type of situation. The results are shown in Fig. 5.2 and compared with the FLAT propensities. It can be noted that the values of the VBW propensities for the residues chosen by us correctly assume the pre-set values.

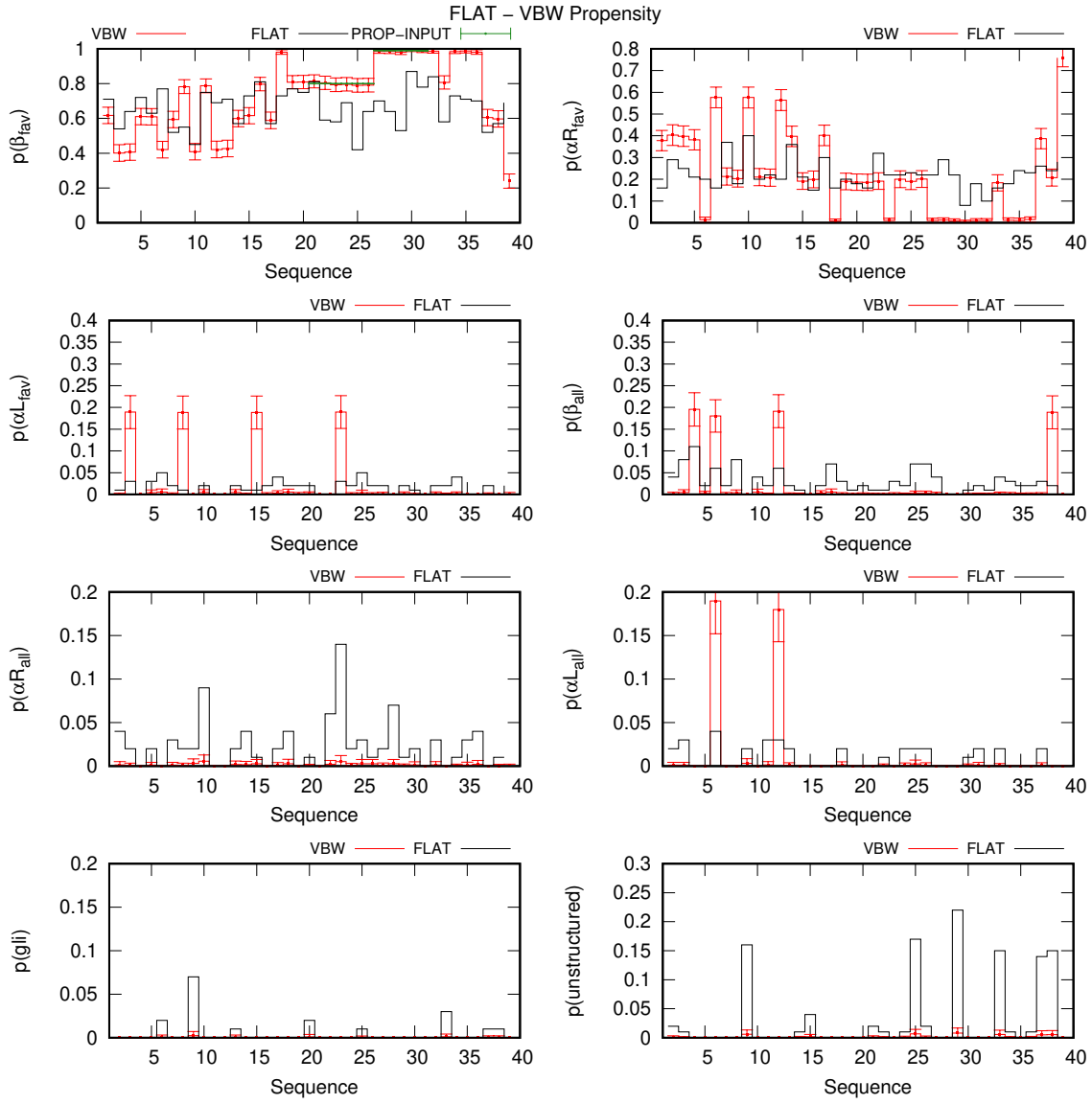


FIGURE 5.2: Known propensity (VBW, red line) imposed for the method's validation in comparison with averaged propensity (FLAT, black line).

After calculating the w_i weights, these were used to simulate the respective scattering curves $I(q, < w_i >)$ (with their relative SDs) thanks to the use of the VBWSAS 1.0 method. The simulated scattering curve are shown in Fig. 5.3:

- The black curve is the scattering curve derived from the w_i described by FLAT propensity shown in Fig. 5.1;
- The red curve is the scattering curve derived from the w_i described by imposed propensity (VBW propensity, Fig. 5.2).

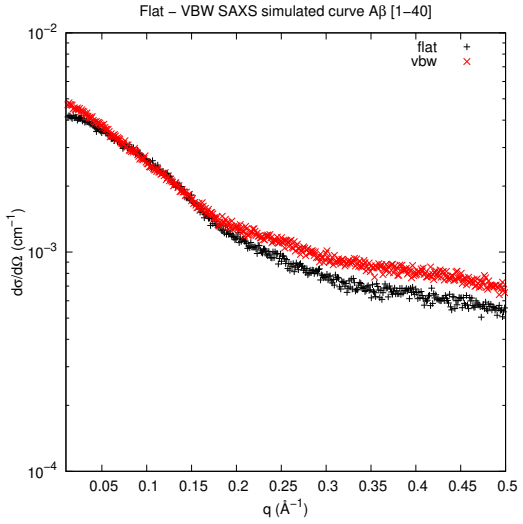


FIGURE 5.3: Simulated SAXS curve for Aβ[1-40]

As you can see from the Fig. 5.3, the scattering curves have significant difference, and it should be noted that a difference in propensity (Fig. 5.2), then translates into an important difference in the scattering curve, confirming how SAS technique can be a good technique for this kind of analysis. Subsequently, the black curve, derived from the propensities imposed by us, was analyzed with the VBWSAS 1.0 method as if it were an experimental data. The method minimizes the α functional, from which the necessary weight factors, called w_i^{sas}, are calculated to calculate the propensities. In Fig. 5.4 is shown the best fitting from VBWSAS 1.0 method.

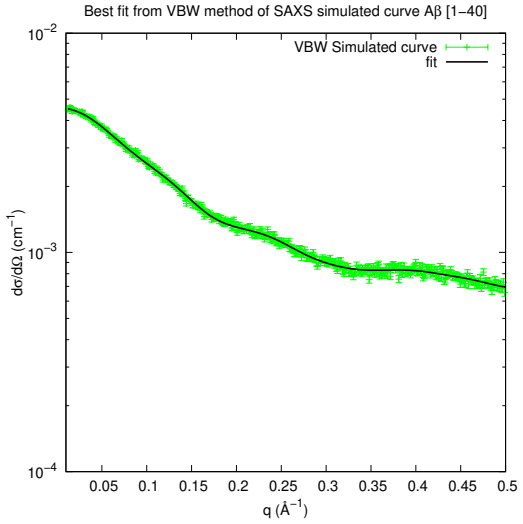


FIGURE 5.4: Best fitting of SAXS curve for Aβ[1-40] calculated with VBWSAS 1.0 method

The Fig. 5.4 shows how the fit is of excellent quality. The χ^2 of the analysis is equal to 1.0. The 100 w_i^{SAS} obtained from VBWSAS 1.0 method can now be used to calculate the propensity of this simulated data as shown in Eq. 4.3 . The output of this last calculation is a plot for each region of the Ramachandran plot where the amino acid sequence is on the abscissa and the propensity value for that particular region on the ordinate. In the Fig. 5.5 the results are shown.

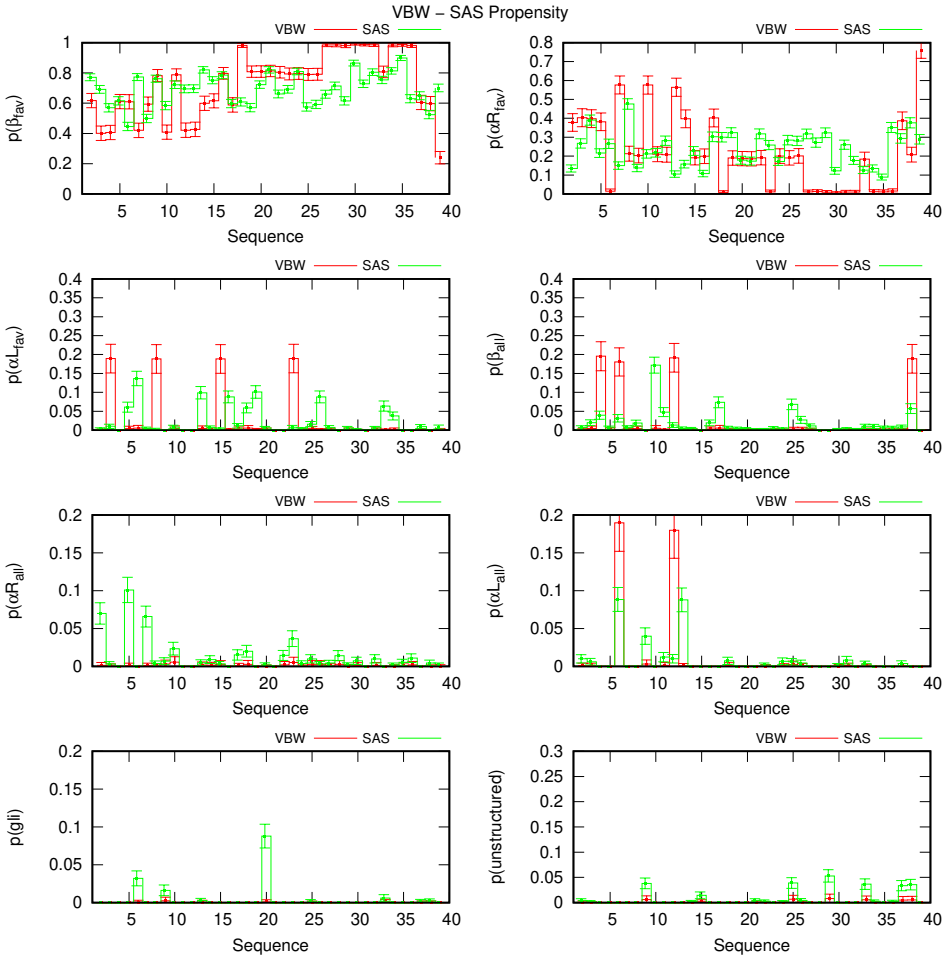


FIGURE 5.5: VBWSAS propensity (SAS, green line) derived from VBWSAS 1.0 method calculation in comparison with Known propensity (VBW, red line) imposed for the method's validation).

The green line in the Fig 5.5 indicates the information obtained from the SAXS curve. The capability of the VBWSAS 1.0 method to obtain information about the propensities is performed by a comparison of the known propensities (red line) with the ones derived by analysis of the simulated SAS curve (green line). The results show that the propensities in the β_{fav} region, in particular those related to the putative nucleation sites (where the propensities were imposed) are quite close, in some cases perfectly overlapped with those previously imposed. This result, considering the large average conformational disorder of the starting ensemble, is very important, because it indicates how the SAS techniques

can be quite sensitive to conformational changes of parts of the amino acid sequence.

5.1.2 αR_{fav} region test

The same analysis was carried out by constraining the propensity in the αR region to the value 0.72 for the region from the aminoacid number 10 to the aminoacid number 15. Also in this case the SAXS curve of each simulated structure obtained by means of the weighted averages on the curves of each structure was generated with the weight factors ,w_i selected by the VBWSAS 1.0 method in order to favor sequences with high α propensities. In Fig. 5.6 is shown the comparison between the FLAT curve of the ensemble (black line) and the VBW curve derived from the new simulated situation (red line).

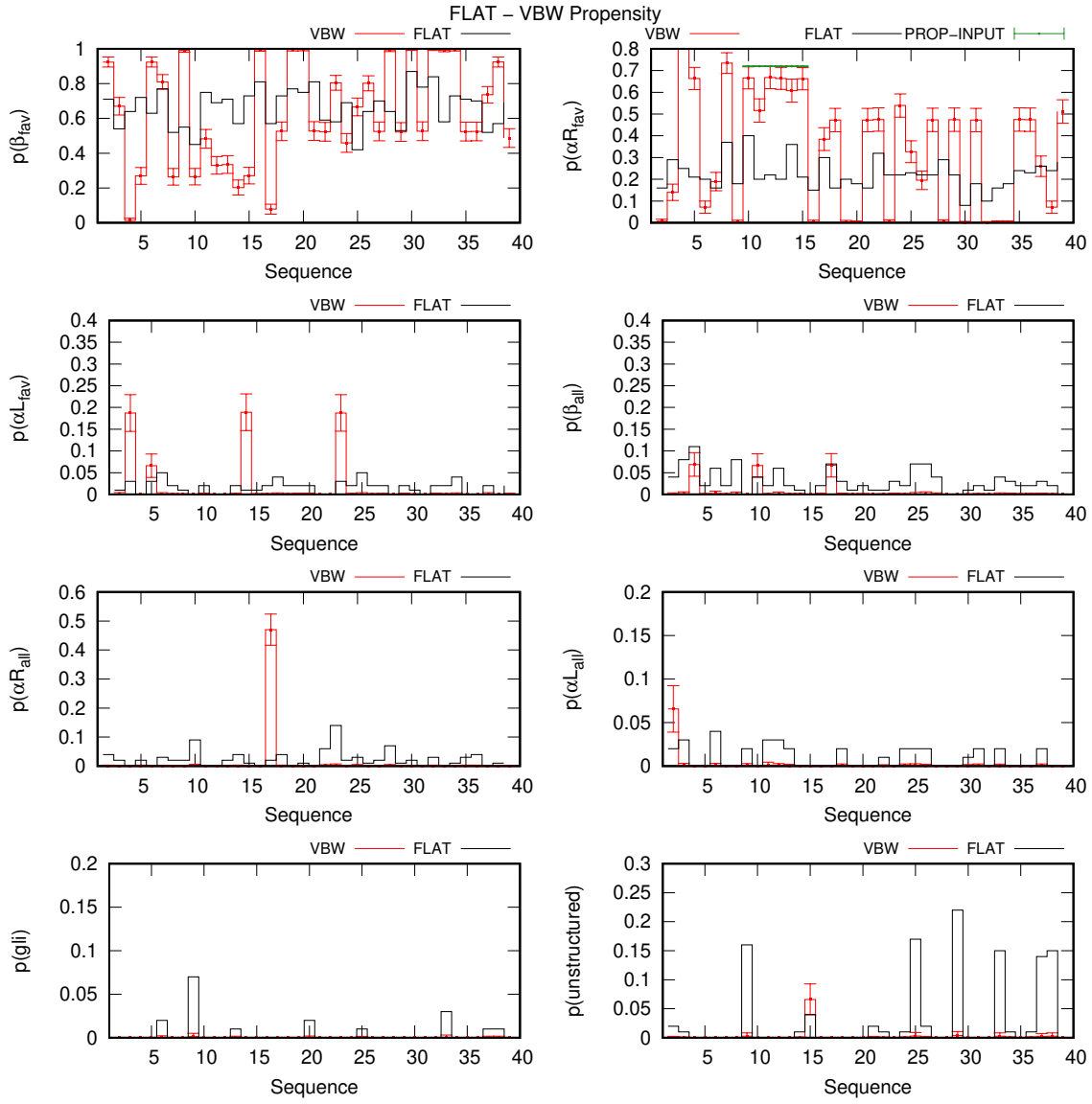


FIGURE 5.6: Known propensity (VBW, red line) imposed for the method's validation in comparison with averaged propensity (FLAT, black line).

The Figure 5.7 shows the simulated SAXS curve calculated with the w_i obtained from VBWSAS 1.0 method (VBW, red line) to simulate an high α propensity, in comparison with the average curve of the ensemble (FLAT curve, black line).

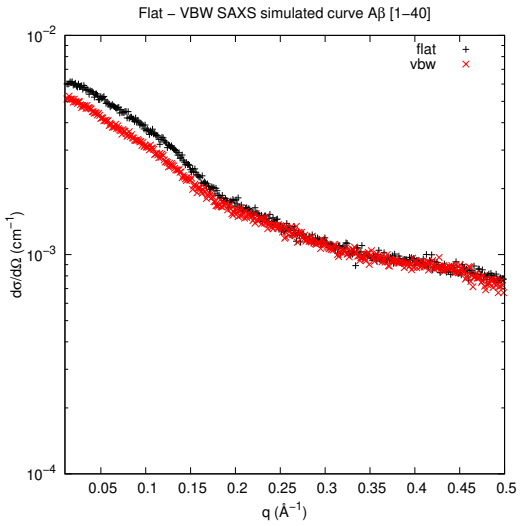


FIGURE 5.7: Simulated SAXS curve for Aβ[1-40]

Subsequently, the analysis of this curve was carried out using the VBWSAS 1.0 method, similarly to what was explained in the previous paragraph. Subsequently, the analysis of this curve was carried out using the VBWSAS 1.0 method, similarly to what was explained in the previous section. The fit of the experimental curve is shown in Fig. 5.8, and confirms an excellent reproducibility of the input curve, which simulates a possible experimental data.

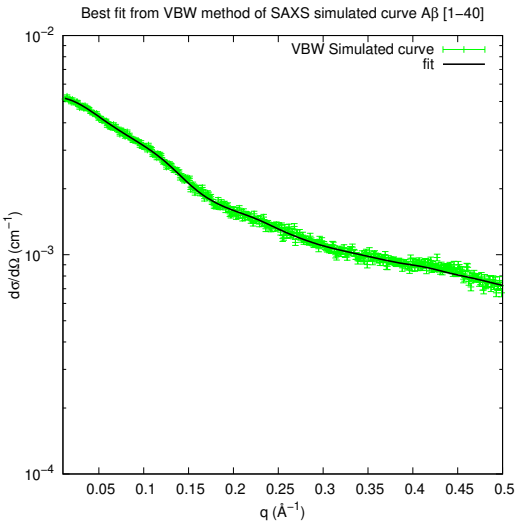


FIGURE 5.8: Best fitting of SAXS curve for Aβ[1-40] calculated with VBWSAS 1.0 method

The results of the propensities are shown in Fig. 5.9 and are very close to the propensities associated to the simulated SAXS experiment, highlighting that the SAXS experiments are suitable to derive conformational propensities.

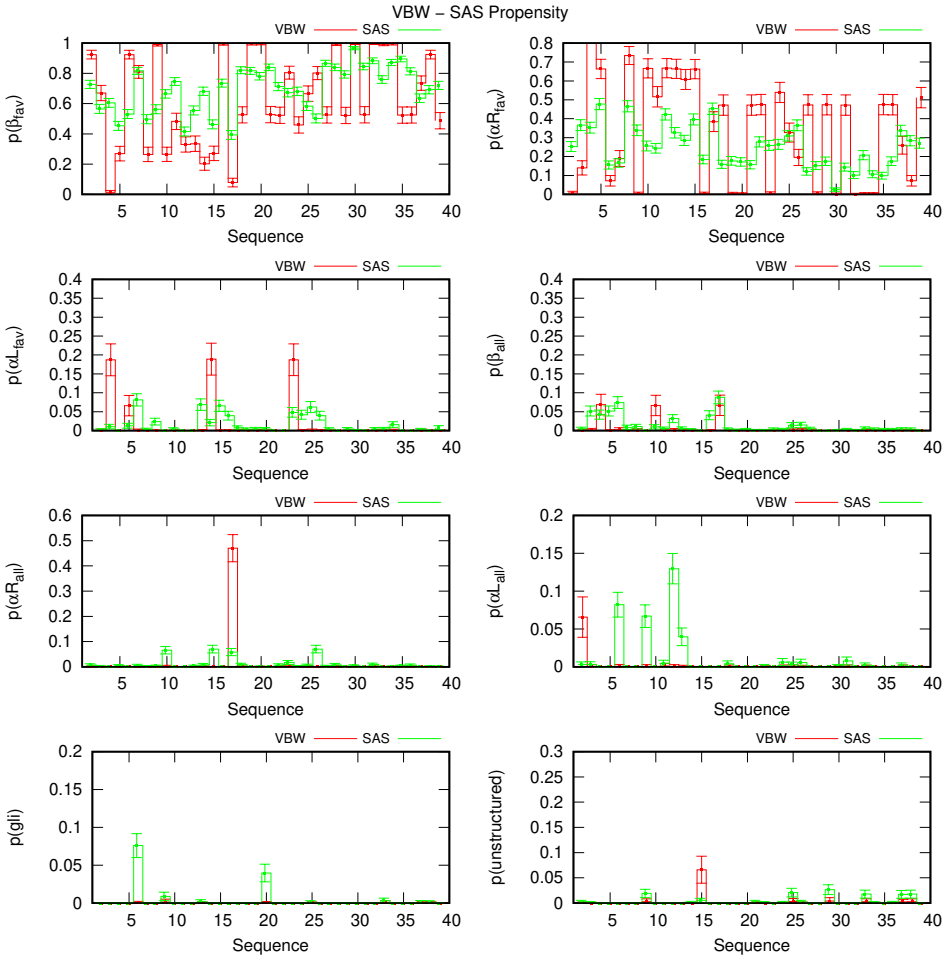


FIGURE 5.9: VBWSAS propensity (SAS, green line) derived from VBWSAS 1.0 method calculation in comparison with Known propensity (VBW, red line) imposed for the method's validation).

The results show firstly how the VBWSAS 1.0 method is sensitive enough to recognize changes in propensity in conformational ensembles, and in addition confirm the usefulness of the SAXS technique in investigating this kind of sample and in acting as a complementary technique to NMR and XRD technique.

Chapter 6

VBW application on experimental data

6.1 Experimental Data

Once the method has been validated by means of simulations, some of which are shown in chapter 5, the VBWSAS method in the versions 1.0, 2.0 and 3.0 has been applied to the experimental data. Hereafter we show the results of the last version of the method, VBWSAS 3.0. As explained in chapter 5, the experimental data have been recorded on the SAXS beamline 29 (BM29) of ESRF, The European Synchrotron in Grenoble. The various experimental sessions were aimed at the study of α - syn WT and some of the most important mutants, reported in the literature, to be the cause of early Parkinson. The mutants analyzed were as follow:

- ALA 30 PRO (A30P);
- ALA 53 TYR (A53T);
- GLU 46 LYS (E46K);
- GLY 51 ASP (G51D).

The experiments were performed, for the WT's protein and for each type of mutant, at different concentrations and at three temperatures (25, 37 and 45°C). The pH of the measured samples was the physiological one. The main objective of the experiment was to obtain the conformational change caused by the various point mutations from the recorded SAXS curves.

6.2 VBWSAS fitting data

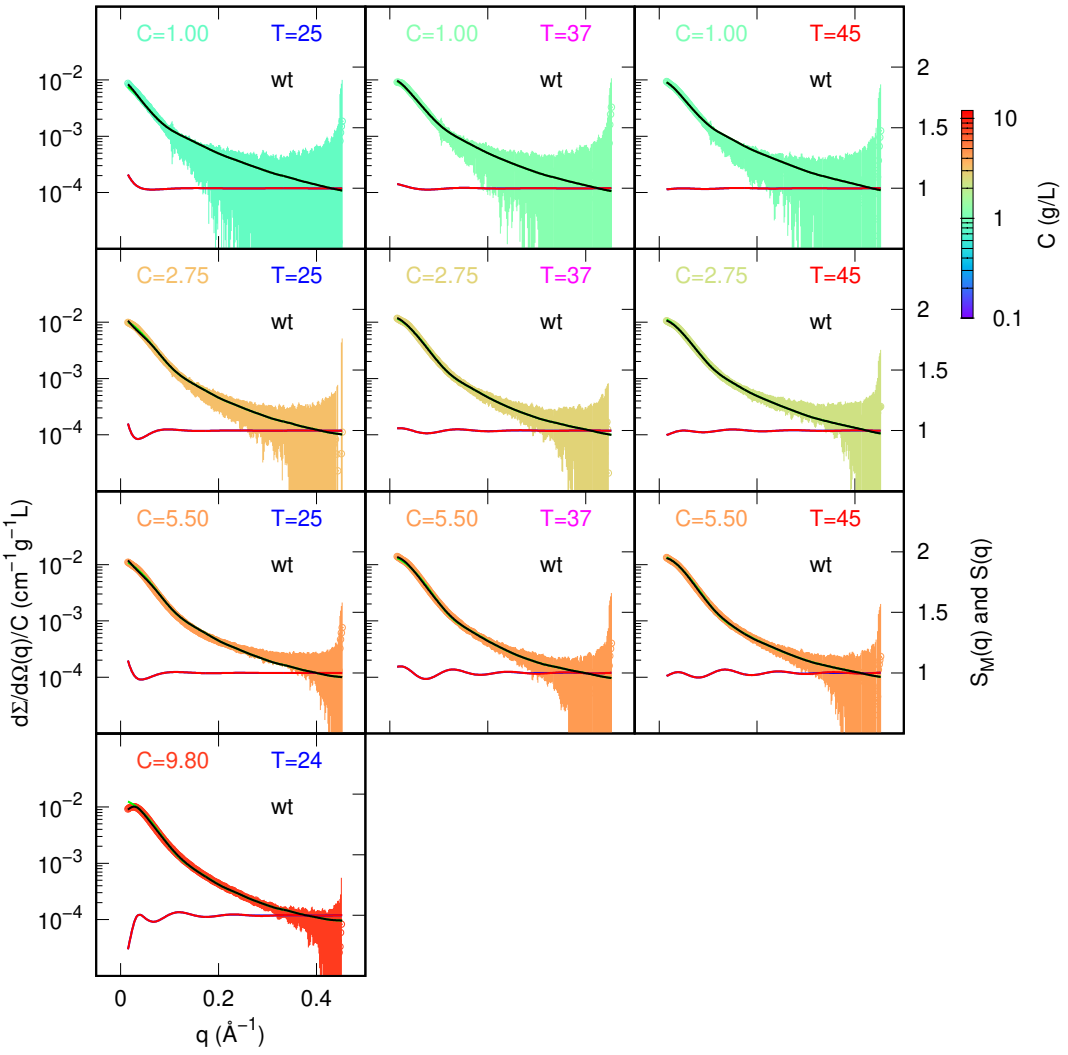
The curves recorded in the European Synchrotron were analyzed with the new VBWSAS 3.0 method. Global fits have been made for each type of protein analyzed in order to make the analysis more reliable:

- First fit WT: 10 curves;
- Second fit G51D: 17 curves;
- Third fit E46K: 11 curves;
- Fourth fit A53T: 18 curves;

- Fifth fit A30P: 4 curves.

All the fitting curves are shown in Fig. 6.1 6.2 6.3 6.4 6.5.

Results, taken together, indicate that the VBWSAS 3.0 method is robust and is able to obtain good fitting results in different composition scenarios.


C=1.00 T=25 wt
C=1.00 T=37 wt
C=1.00 T=45 wt
C=2.75 T=25 wt
C=2.75 T=37 wt
C=2.75 T=45 wt
C=5.50 T=25 wt
C=5.50 T=37 wt
C=5.50 T=45 wt
C=9.80 T=24 wt
dΣ/dΩ(q)/C (cm^{-1}g^{-1}L)
q (Å^{-1})
S$_M$(q) and S(q)
C (g/L)
10
1
0.1

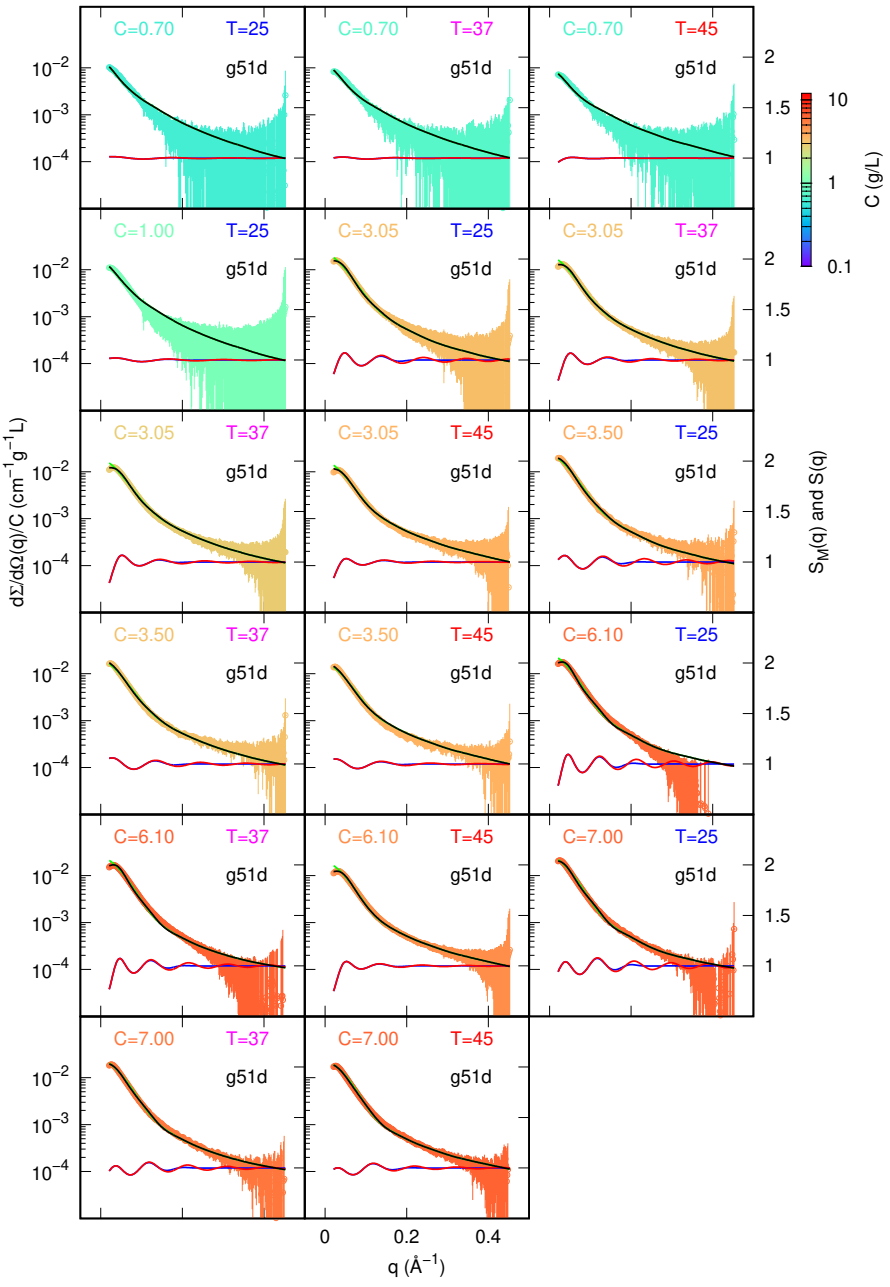


FIGURE 6.2: Fitting curve of the global fitting on G51D α-syn's mutant

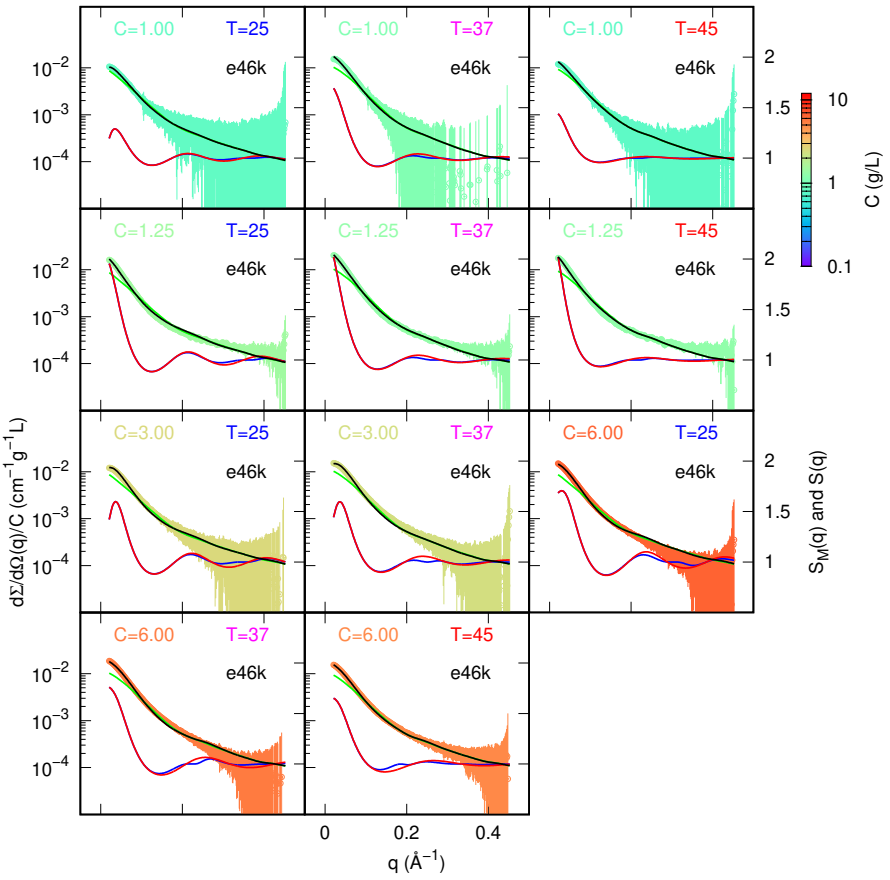


FIGURE 6.3: Fitting curve of the global fitting on E46K α-syn's mutant

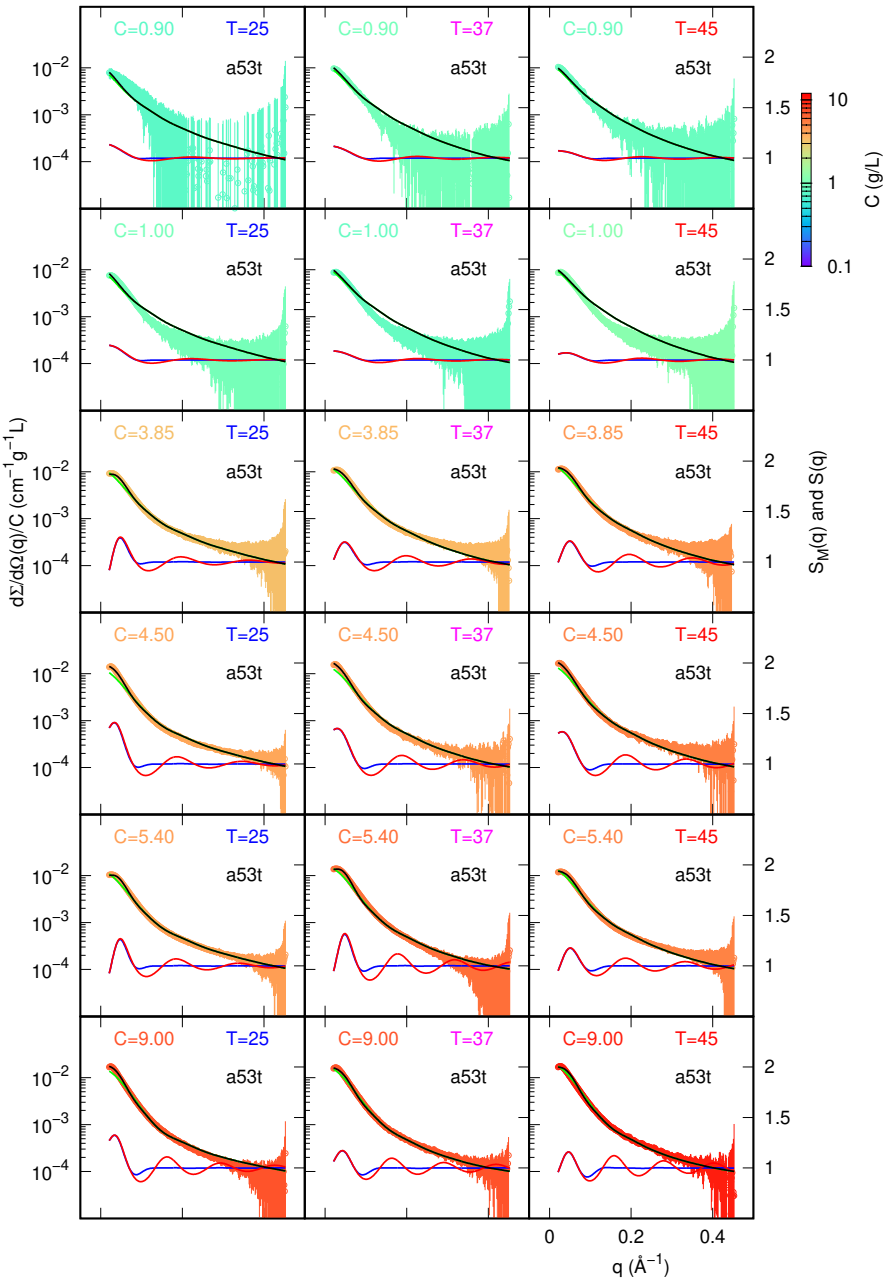


FIGURE 6.4: Fitting curve of the global fitting on A53T *alpha*-syn's mutant

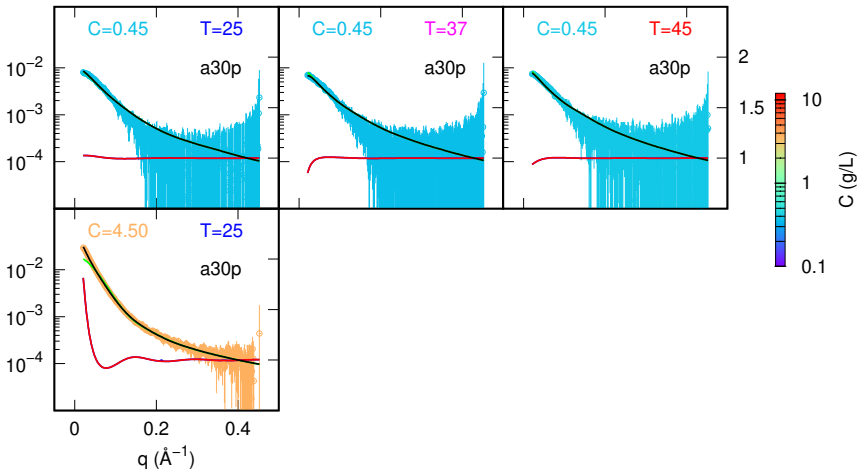


FIGURE 6.5: Fitting curve of the global fitting on A30P α-syn's mutant

Figures 6.1, 6.2, 6.3, 6.4 and 6.5 shows all the experimental curves normalized by concentration and colored according to the concentration of the protein in the sample. The best fit curves obtained from the analysis with the VBWSAS 3.0 method are shown in black. The structural factor, $S(q)$ calculated by the VBWSAS 3.0 method is shown with a red line. In each panel the following information are shown:

- the concentration of the sample;
- the temperature at which the experiment was performed;
- the type of protein under examination (WT or mutant).

The results derived from the VBWSAS analysis on the curves shown are:

- propensity;
- correlation maps of weight factors w_i;
- thermodynamic parameters (ΔH^0, ΔS, ΔCp) describing how the system changes thermodynamically as a function of the mutation;

These results are shown and discussed in the next section 6.3.

6.3 Results and Discussion

6.3.1 Propensity

In this section the propensities derived from VBWSAS analysis are shown. Since the VBWSAS 3.0 method is able to determine the variation with concentration and temperature of the weight of the conformers (Eq. 4.48), all the results here shown have been calculated at three representative concentrations 2, 5 and 10 g/L and three temperatures 25, 37 and 45°C. In all the figures below, the difference between the propensity of each mutant and the one of the wild type are shown. This is an optimum representation, able to to highlight how much the mutant behaves differently from the WT. The graphs considered the variation of propensity to be in a certain sequence only if it maintain the same sign (positive or negative) for at least 5 subsequent amino acids. For this reason, in the graphs there are parts of the sequence that do not show any value. Indeed, the lack of result in certain portion of the graph is simply the result of very oscillating variation of propensities in that part of the sequence. In this way we are able to highlight the parts of sequences that are more likely to populate a given region, in accordance with secondary structure prediction methods.

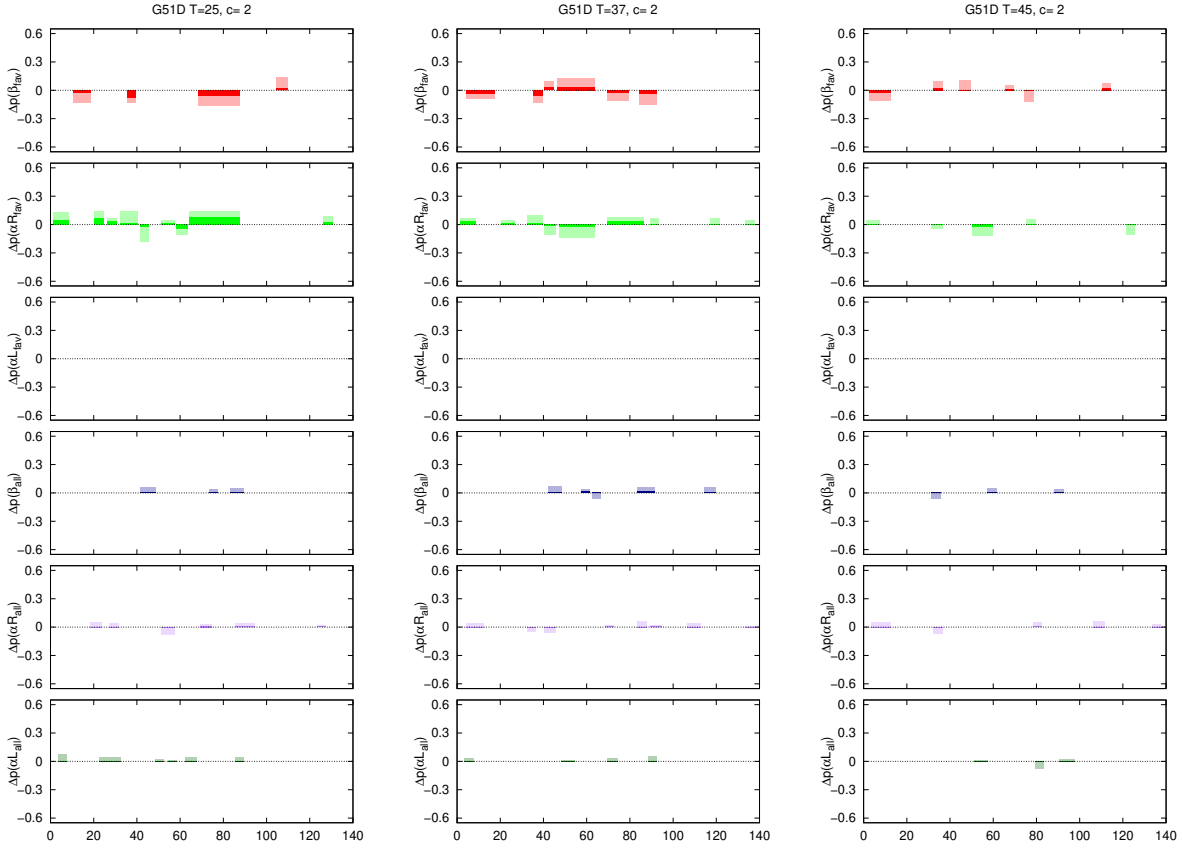


FIGURE 6.6: Propensity G51D in comparison with WT α-synuclein. C=2 g/L

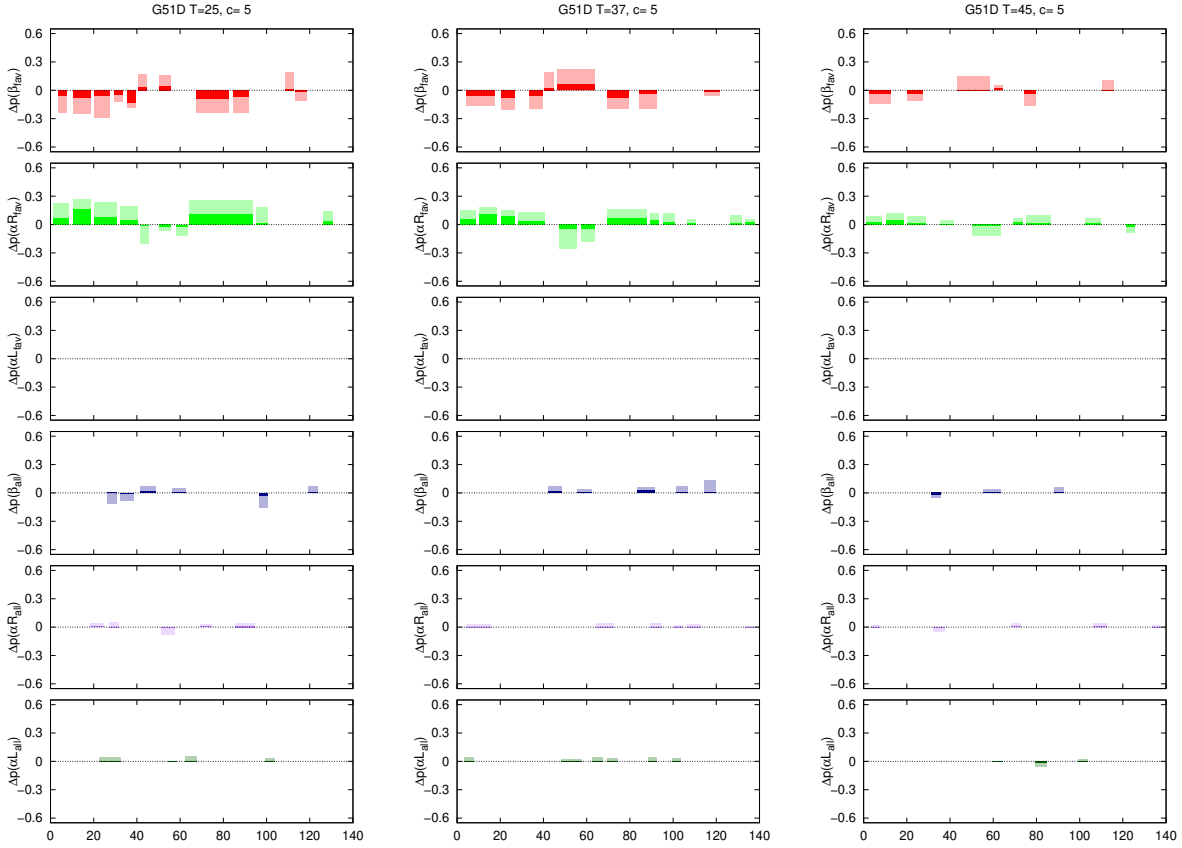


FIGURE 6.7: Propensity G51D in comparison with WT α-synuclein. C=5 g/L

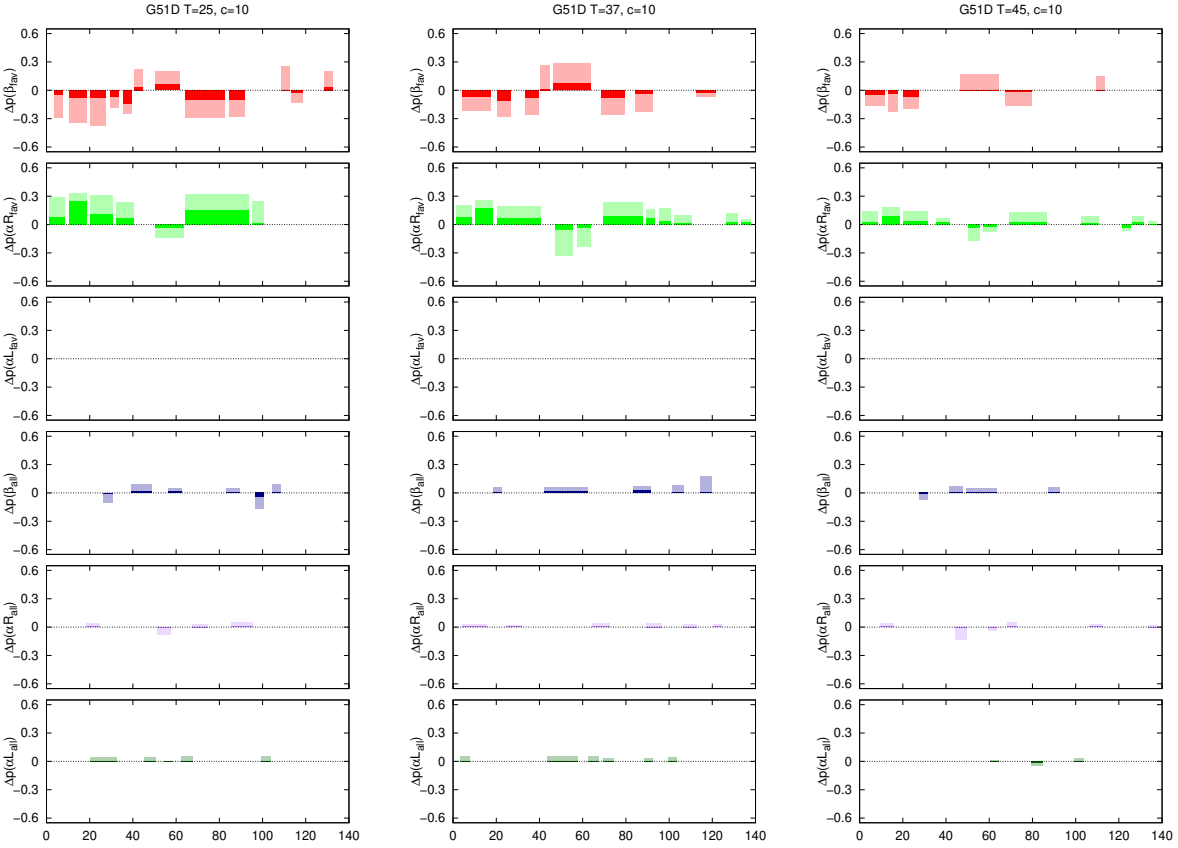


FIGURE 6.8: Propensity G51D in comparison with WT α-synuclein. C=10 g/L

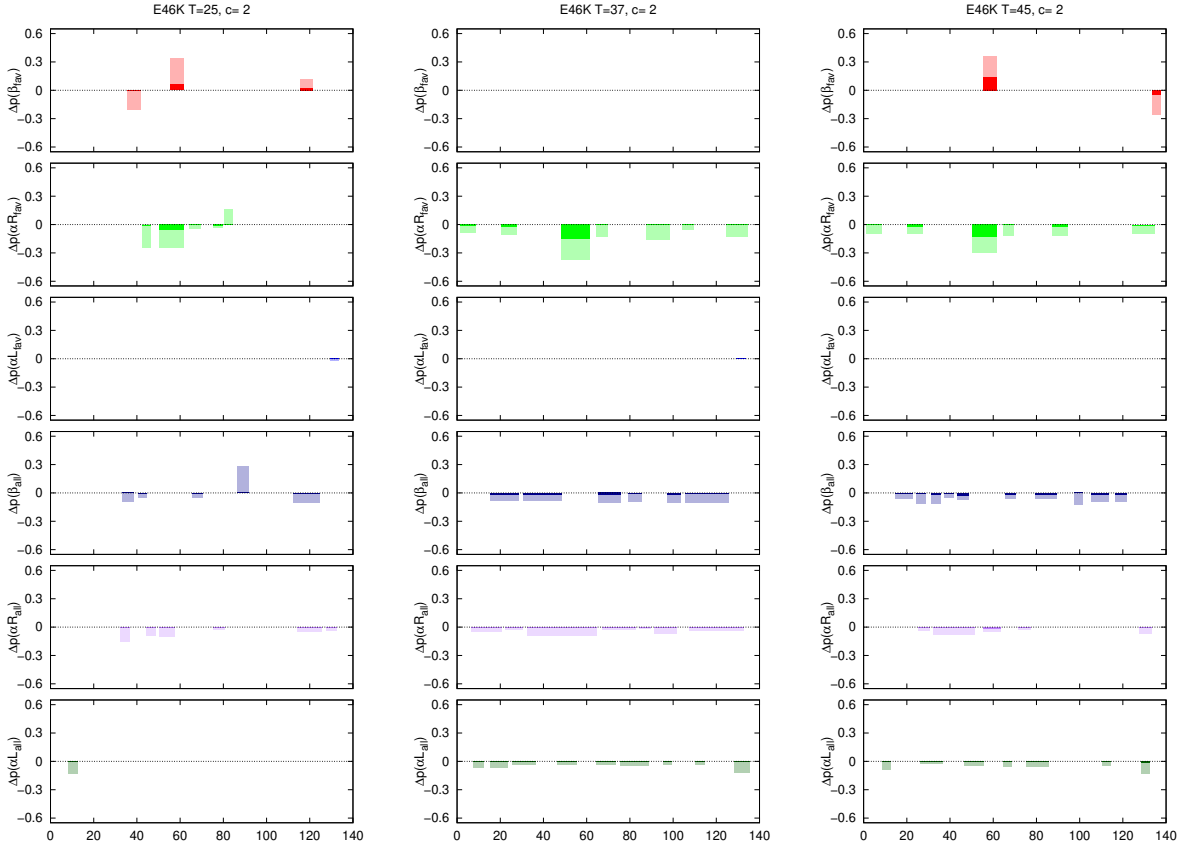


FIGURE 6.9: Propensity E46K in comparison with WT α-synuclein. C=2 g/L

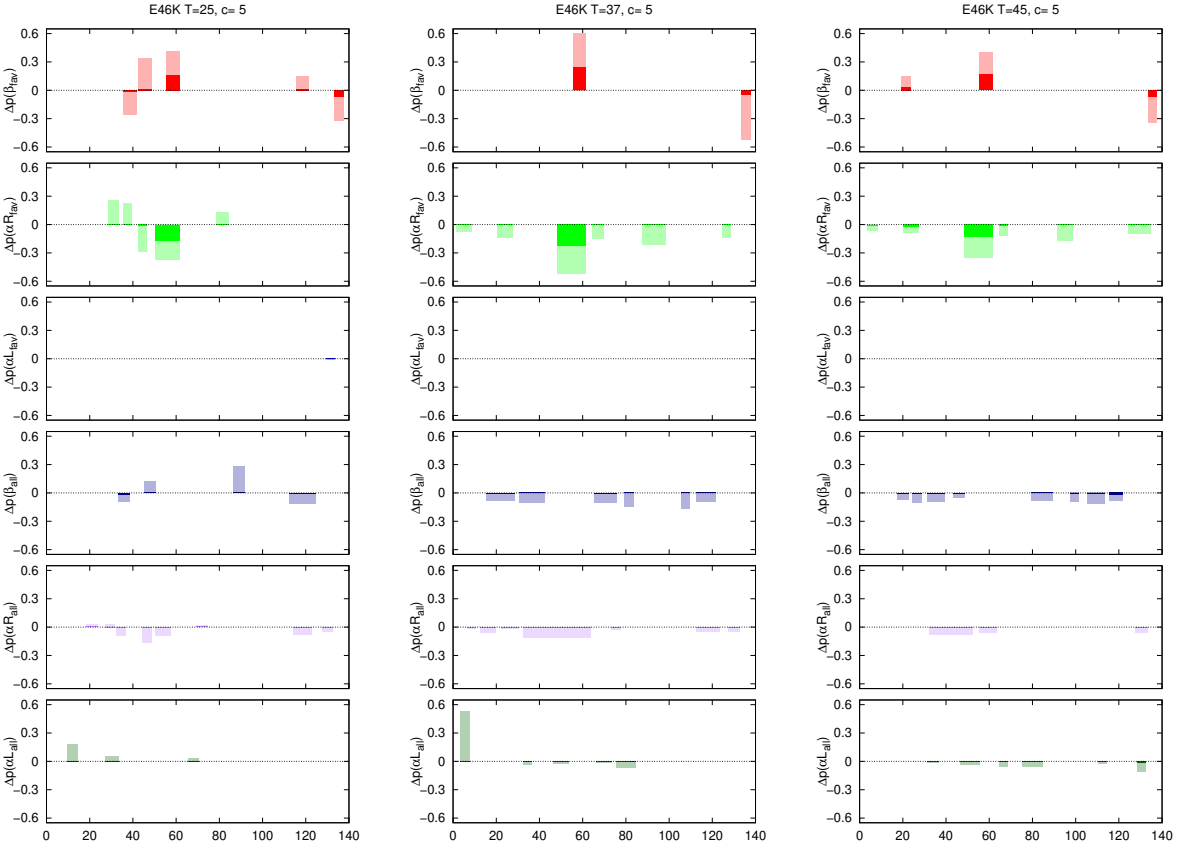


FIGURE 6.10: Propensity E46K in comparison with WT α-synuclein. C=5 g/L

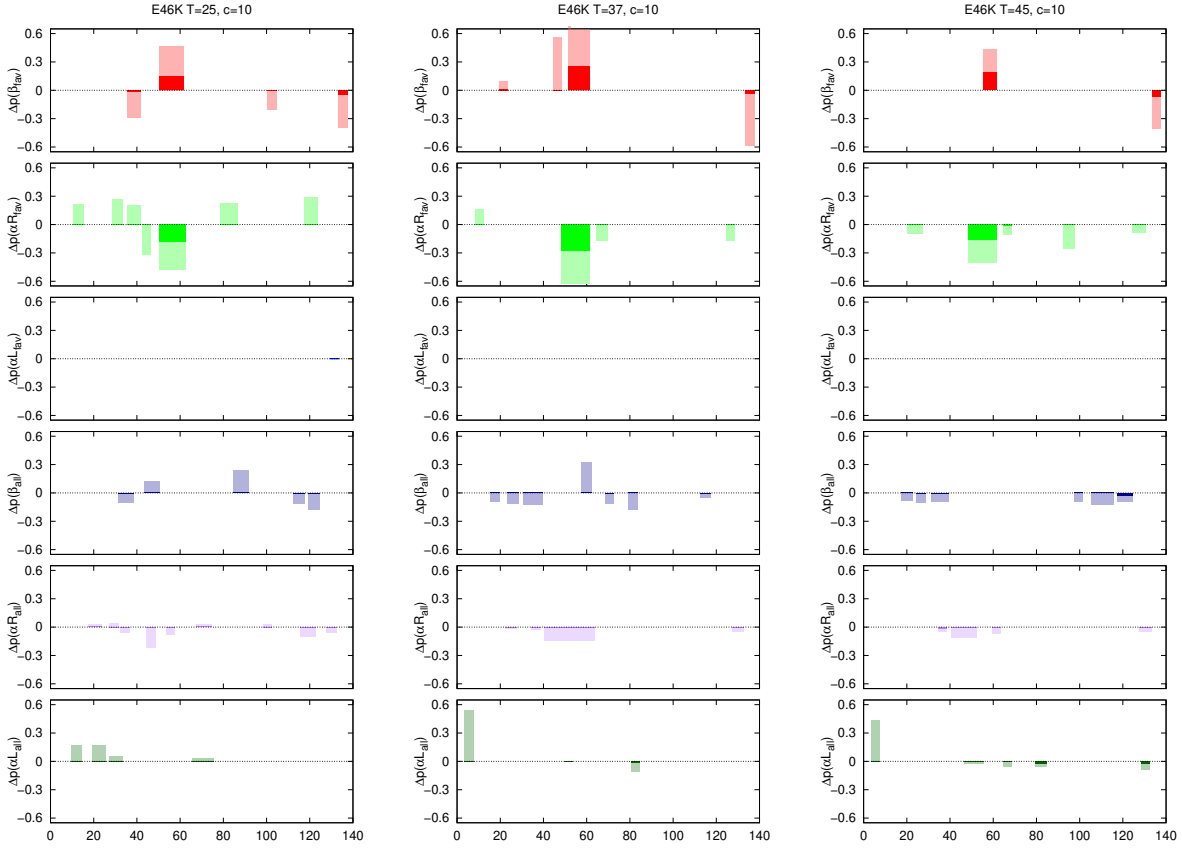


FIGURE 6.11: Propensity E46K in comparison with WT α-synuclein. C=10 g/L

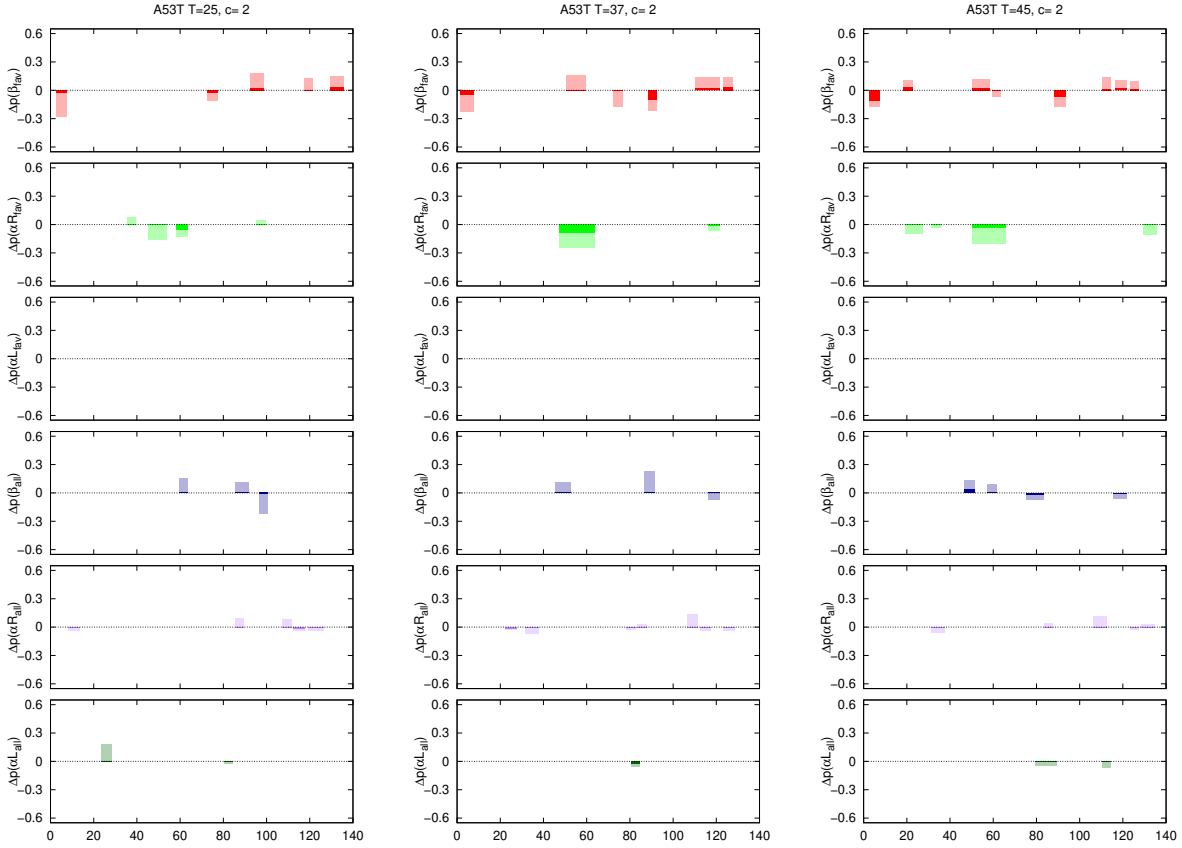


FIGURE 6.12: Propensity A53T in comparison with WT α-synuclein. C=2 g/L

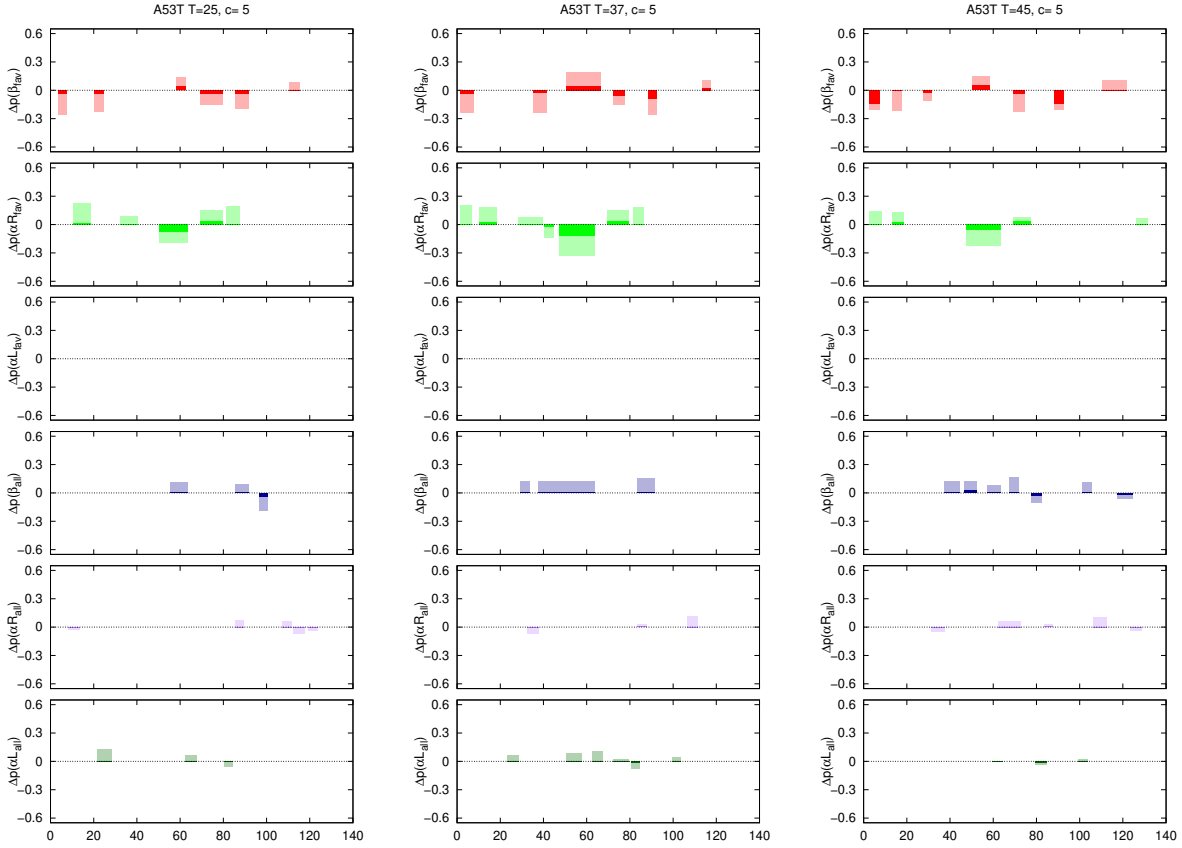


FIGURE 6.13: Propensity A53T in comparison with WT α-synuclein. C=5 g/L

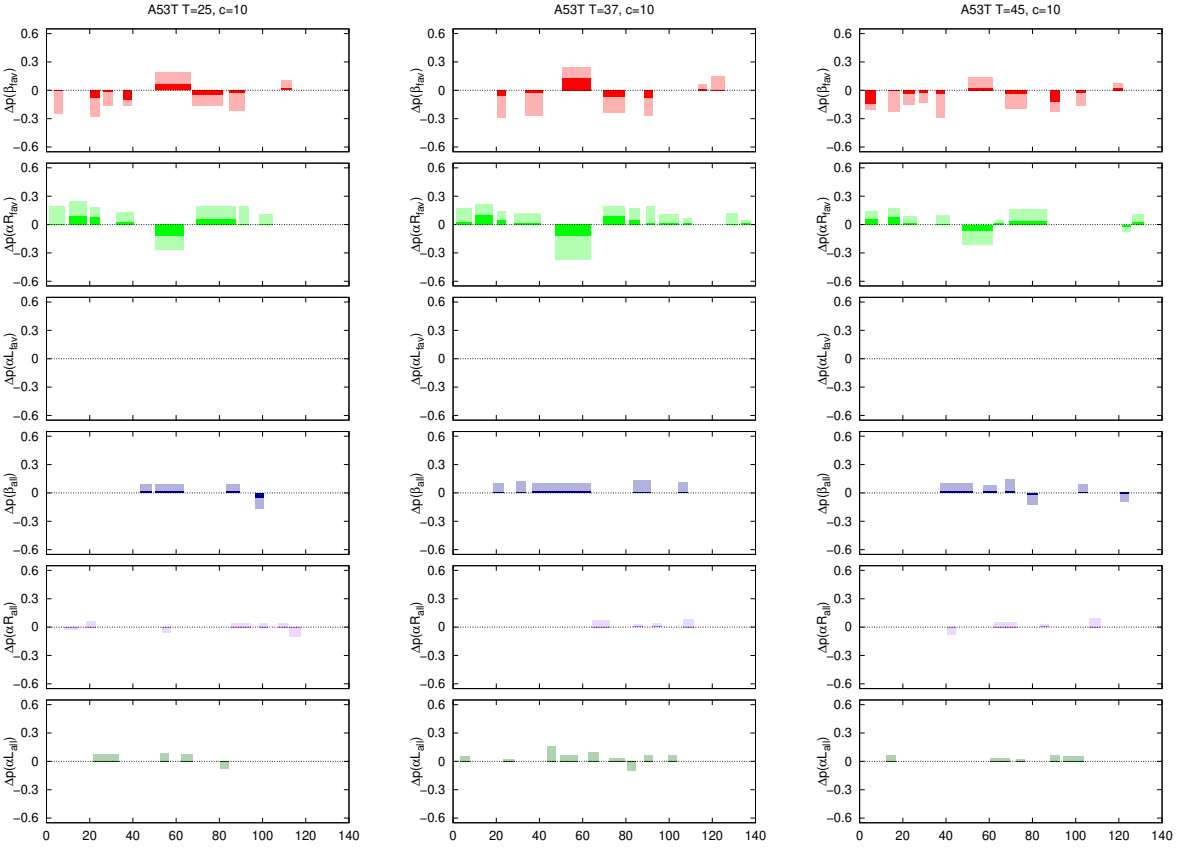


FIGURE 6.14: Propensity A53T in comparison with WT α-synuclein. C=10 g/L

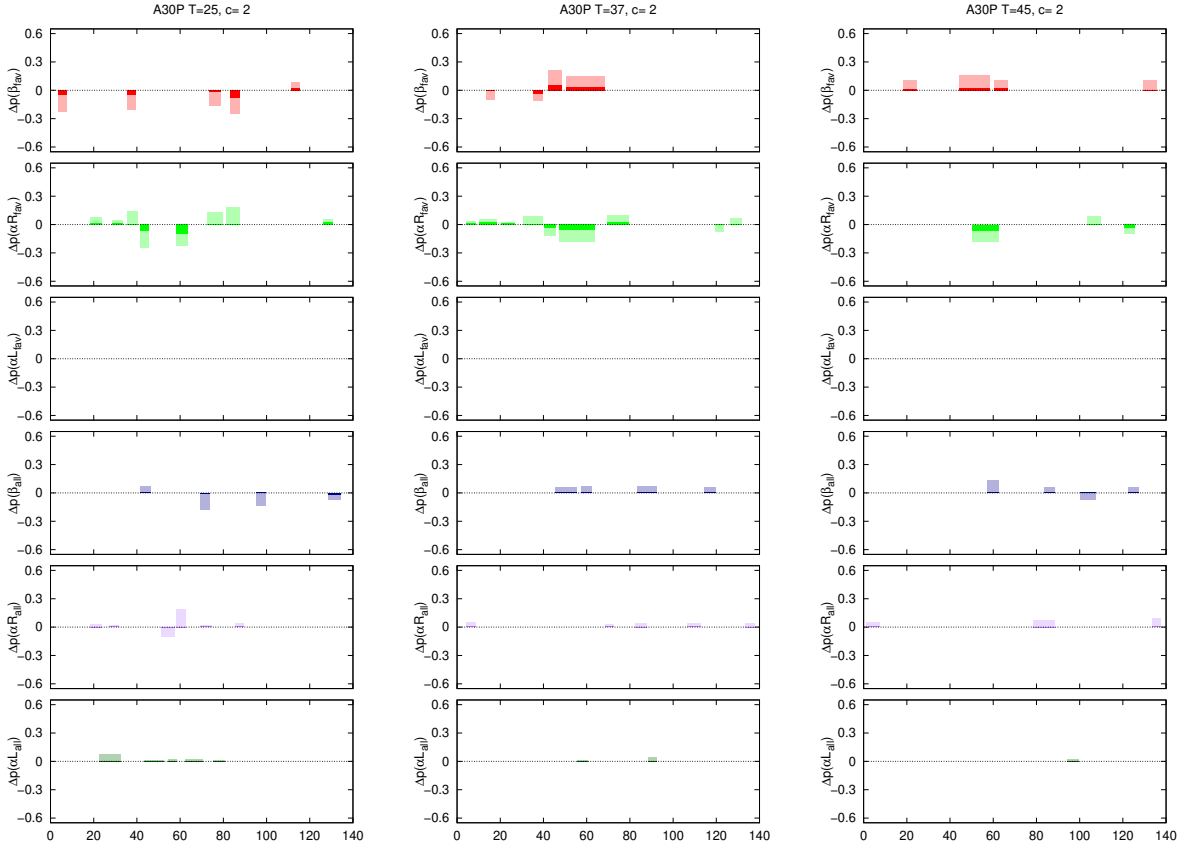


FIGURE 6.15: Propensity A30P in comparison with WT α-synuclein. C=2 g/L

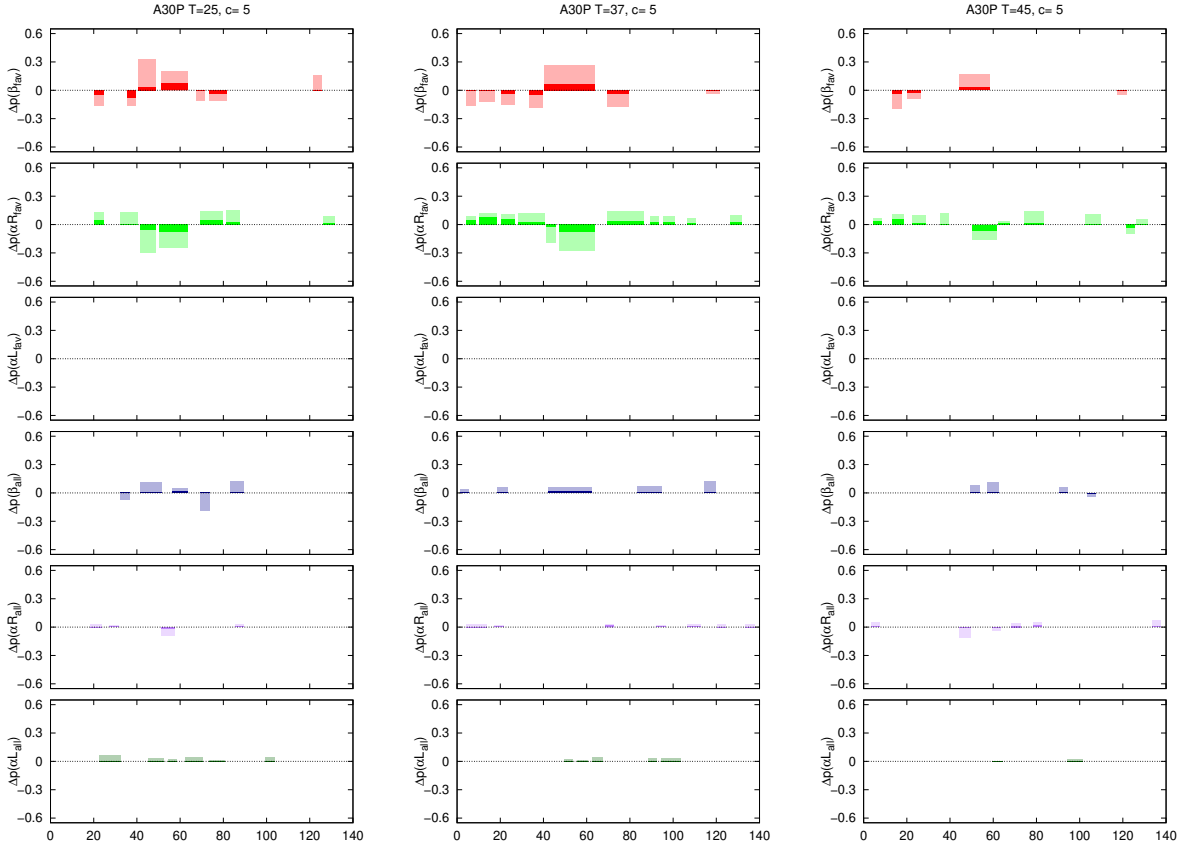


FIGURE 6.16: Propensity A30P in comparison with WT α-synuclein. C=5 g/L

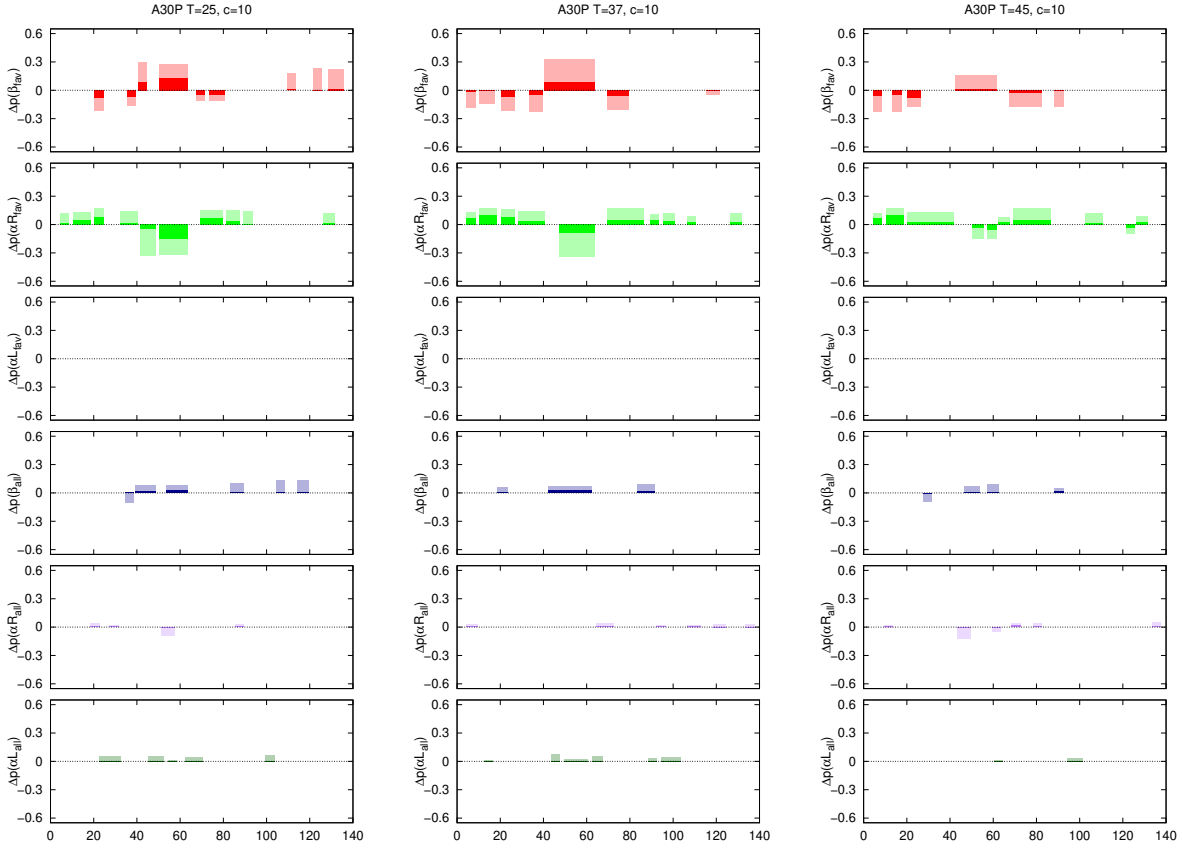


FIGURE 6.17: Propensity A30P in comparison with WT α-synuclein. C=10 g/L

6.3.2 Temperature's Effect

As shown in [27] the temperature has the effect of increasing the aggregation propensity in β - structures of the A53T and A30P mutants (Fig. 6.18). Looking at the graphs of propensity, we can confirm this effect. Up to 37 degrees there is an increase in the propensity in the A53T and A30P mutants, after this temperature it goes back to a more disordered situation.

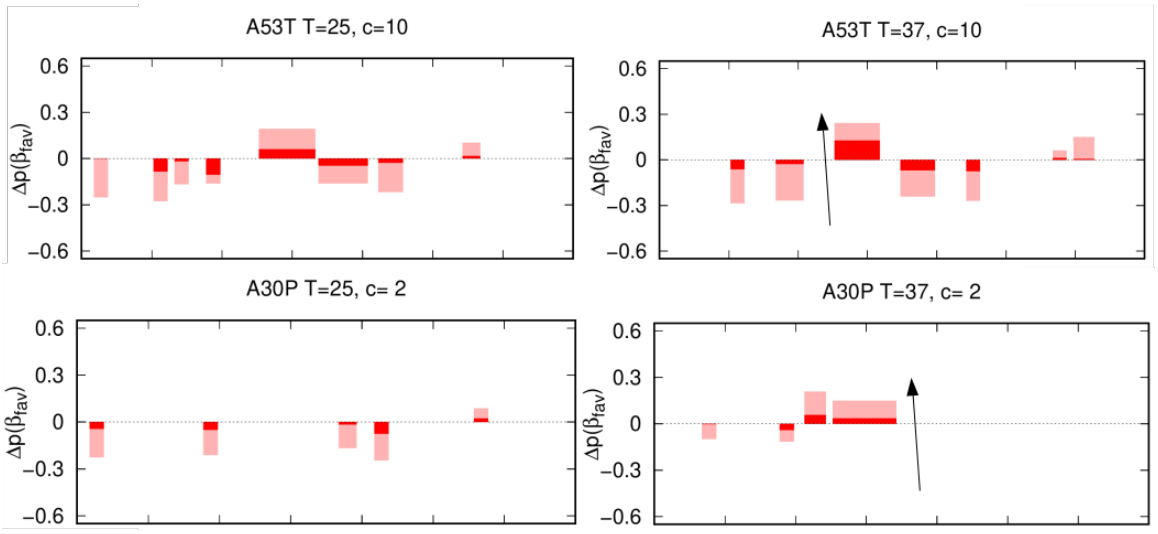


FIGURE 6.18: Increased β propensity as a function of temperature for A53T and A30P mutants.

We found the same temperature's effect also for the E46K mutant (Fig. 6.19) , while for the G51D mutant the results are discordant.

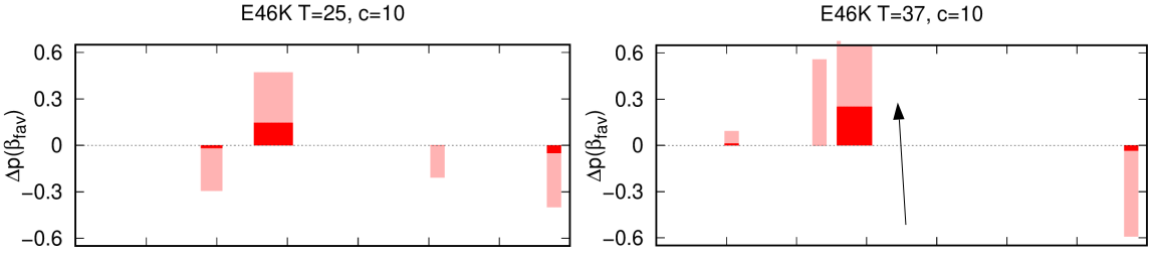


FIGURE 6.19: Increased β propensity as a function of temperature for E46K mutant.

6.3.3 Concentration's Effect

Analyzing the propensities with attention to concentration it is possible to notice how, depending on the concentration of protein in solution, there is an increase in β - propensity as reported in [27]. In the Fig. 6.20 I report the most evident trends as a function of the concentration of mutants A53T and E46K, but this dependence of β - propensity as a function of the concentration is valid for all four mutants.

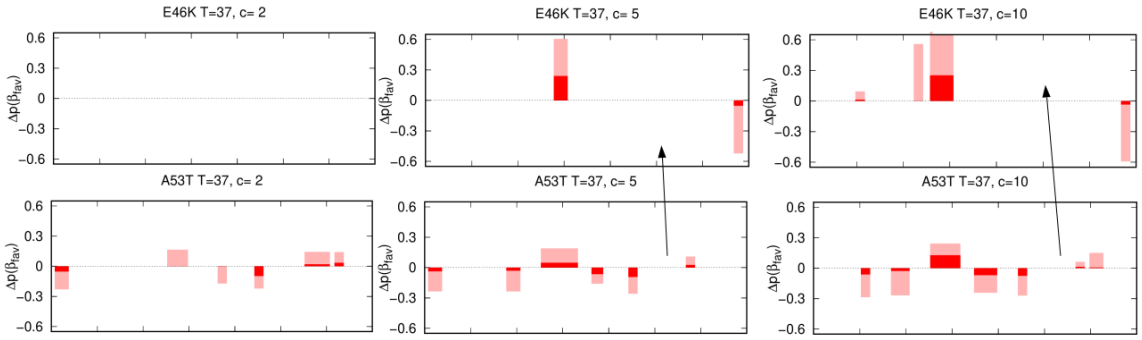


FIGURE 6.20: Increased β propensity as a function of protein's concentration for A53T and E46K mutants.

6.3.4 Conformational's changes

The part that seems to be more involved in the phenomena of aggregation is the N - terminus of the protein. Looking at all the graphs in the large panel of the figures of propensities, it is possible to notice how on average all the mutants show an increase in β - structure compared to the WT protein. This result is in agreement with what has already been confirmed for what concerns the mutants A30P and A53T [27], but from our analysis this turns out to be valid also for the E46K and G51D mutants.

Furthermore, by analyzing the E46K mutant more carefully (Fig. 6.21), it is to be noted

that, unlike the other mutants almost exclusive conformational change of the region in which the point mutation took place. This result is in agreement with [28] and increases the robustness of our analysis, confirming the sensitivity of the method in detecting conformational changes.

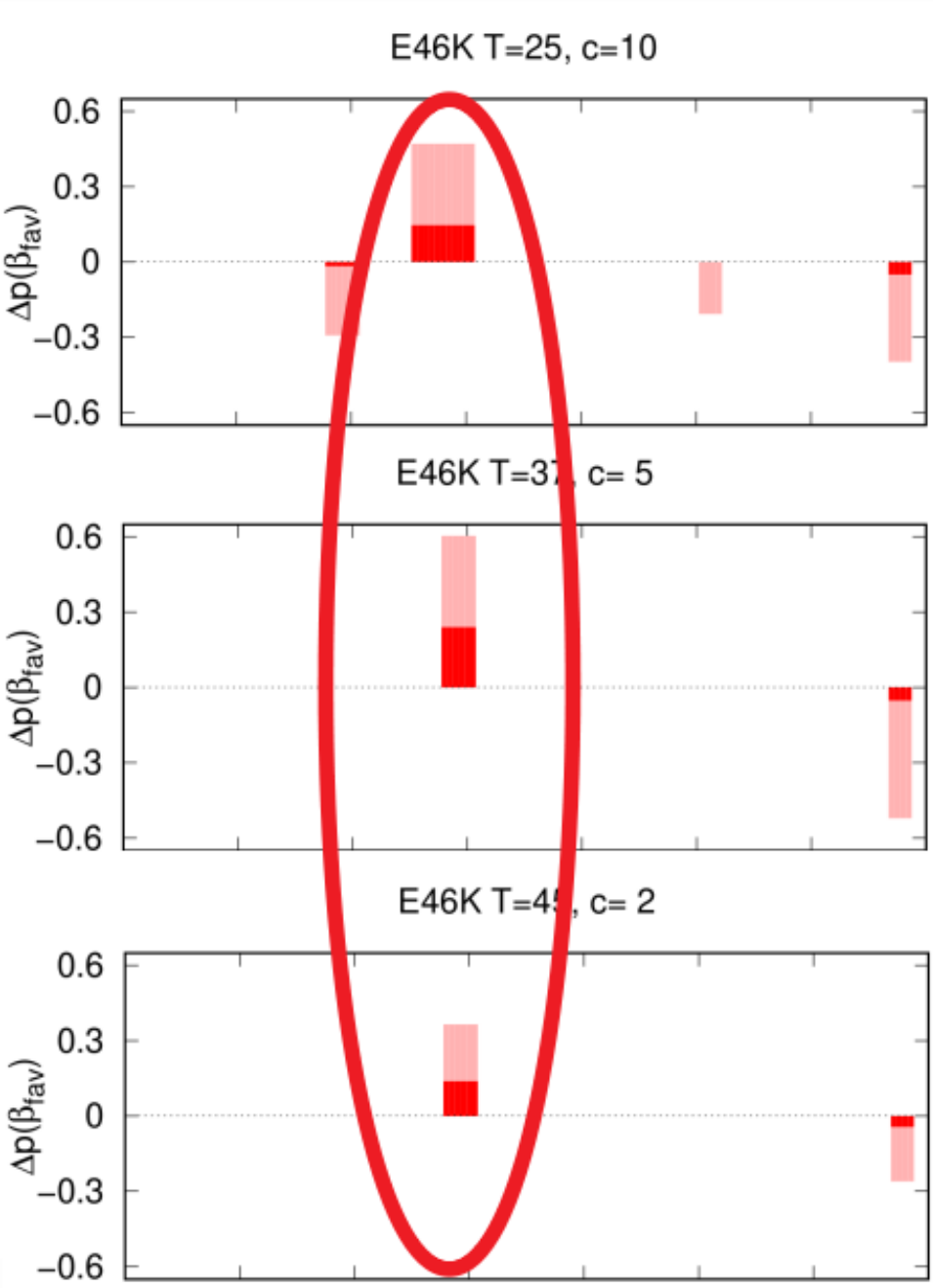


FIGURE 6.21: The point mutation reflects a conformational change in a specific region of the sequence in the E46K mutant, while leaving the remaining part unchanged.

It is known that all mutants have a greater aggregation rate than the wild type [27, 28]. This aspect is difficult to evaluate, but through the VBWSAS analysis it is possible to calculate the structural factors, which measure the protein-protein interaction present in the sample. Increased protein-protein interaction is likely to explain a higher aggregation rate. In fact, looking at the $S(q)$ (Fig. 6.1 6.2 6.3 6.4 6.5) it is possible to see how in the mutants there is much more interaction than in the WT protein. In the future this aspect will be analyzed in more detail.

6.3.5 Correlation's Maps

The weight factors, w_i, elaborated by the VBWSAS 3.0 method are very important because, in addition to being used in the calculation of the propensity, they express the probability of that determined conformation to be populated. To try to better interpret the data, correlation maps have been made, aimed to show which are the most discrimanated conformations between the WT and mutant α-syn. Correlation maps, calculated from the thermodynamic parameters derived by SAXS analysis, are shown in Figs. 6.22-6.25 for three selected values of protein concentration and temperature.

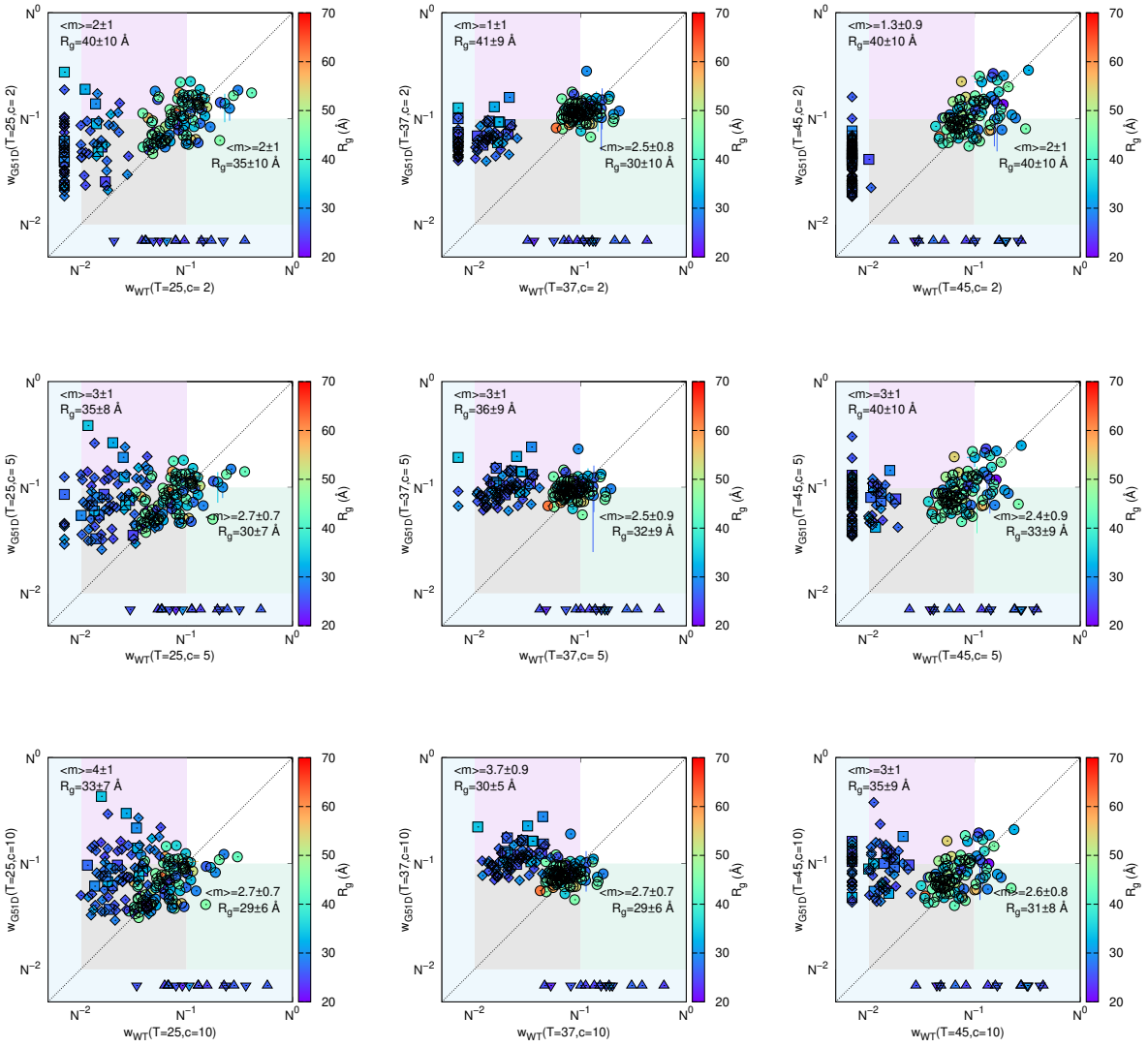


FIGURE 6.22: Correlation map of the weights of the $N = 189$ conformations (Gurry's ensemble) between WT (horizontal axis) and G51D (vertical axis). Each symbol represents the i-conformation and has been color coded on the basis of its R_g, according to the color palette on the right. Symbols are assigned according to Gurry's classification shown in Fig. 4.5. Circles: monomers (A); down-sided triangles: helical-rich trimers (B); up-sided triangles strand-rich trimers (C); squares: helical-rich tetramers (D); diamonds: strand-rich tetramers (E). The purple (light-green) off-diagonal quadrant indicates the group of conformations with w_i simultaneously greater (lower) than N^{-1} for G51D and lower (greater) than N^{-1} for WT. In each off-diagonal quadrant the average values of R_g and $<m>$ of the group of conformations present in the quadrant are reported. Conformations with $w_i < N^{-2}$ for G51D and WT are shown in the middle of the left and bottom cyan strips, respectively. Rows from top to bottom refer to $c = 2$, 5 and 10 g/L. Columns from left to right refer to $T = 25$, 37 and 45°C.

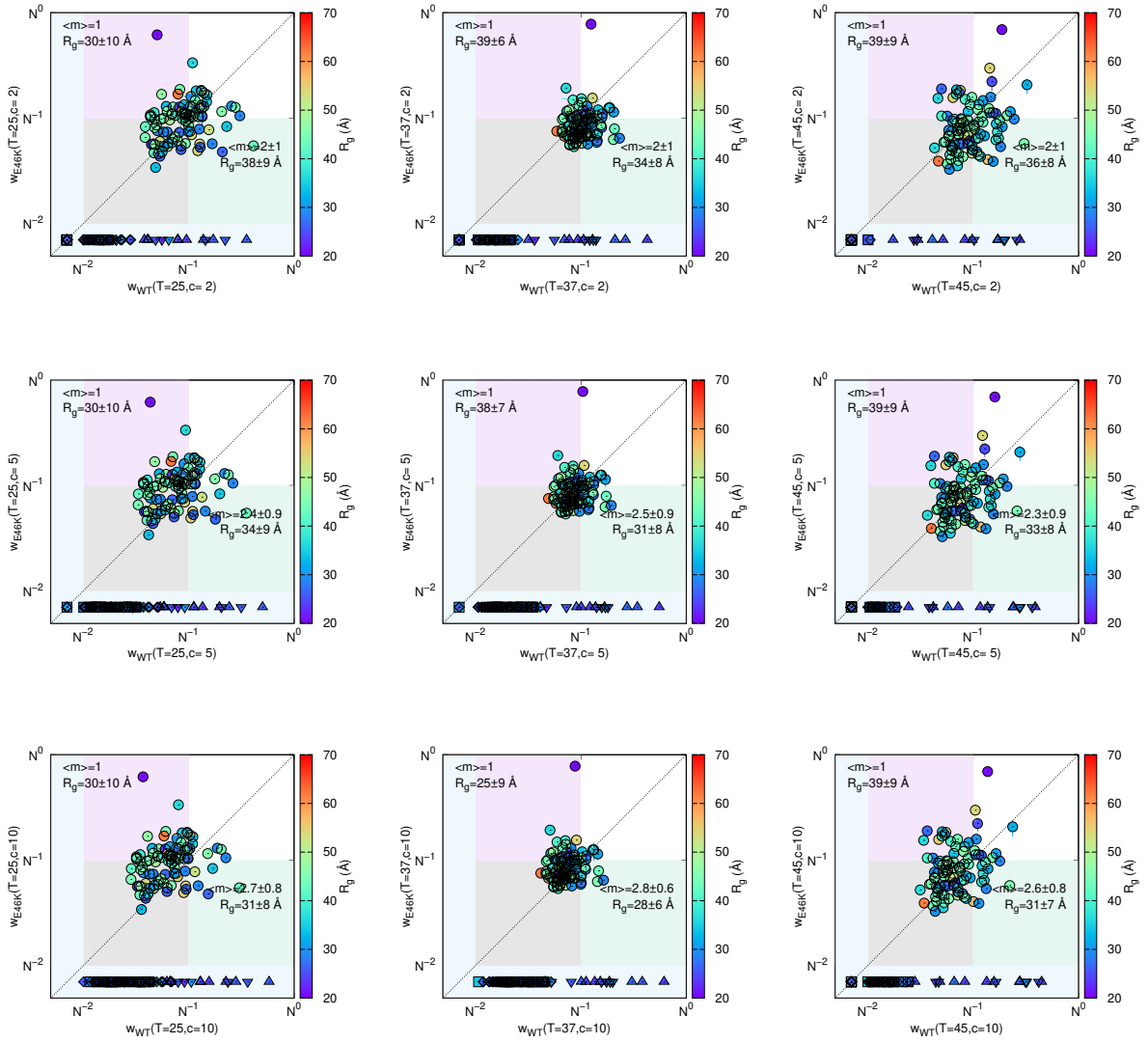


FIGURE 6.23: Correlation map of the weights of the $N = 189$ conformations (Gurry's ensemble) between WT (horizontal axis) and E46K (vertical axis). Each symbol represents the i-conformation and has been color coded on the basis of its R_g, according to the color palette on the right. Symbols are assigned according to Gurry's classification shown in Fig. 4.5. Circles: monomers (A); down-sided triangles: helical-rich trimers (B); up-sided triangles strand-rich trimers (C); squares: helical-rich tetramers (D); diamonds: strand-rich tetramers (E). The purple (light-green) off-diagonal quadrant indicates the group of conformations with w_i simultaneously greater (lower) than N^{-1} for E46K and lower (greater) than N^{-1} for WT. In each off-diagonal quadrant the average values of R_g and $<m>$ of the group of conformations present in the quadrant are reported. Conformations with $w_i < N^{-2}$ for E46K and WT are shown in the middle of the left and bottom cyan strips, respectively. Rows from top to bottom refer to $c = 2$, 5 and 10 g/L. Columns from left to right refer to $T = 25$, 37 and 45°C.

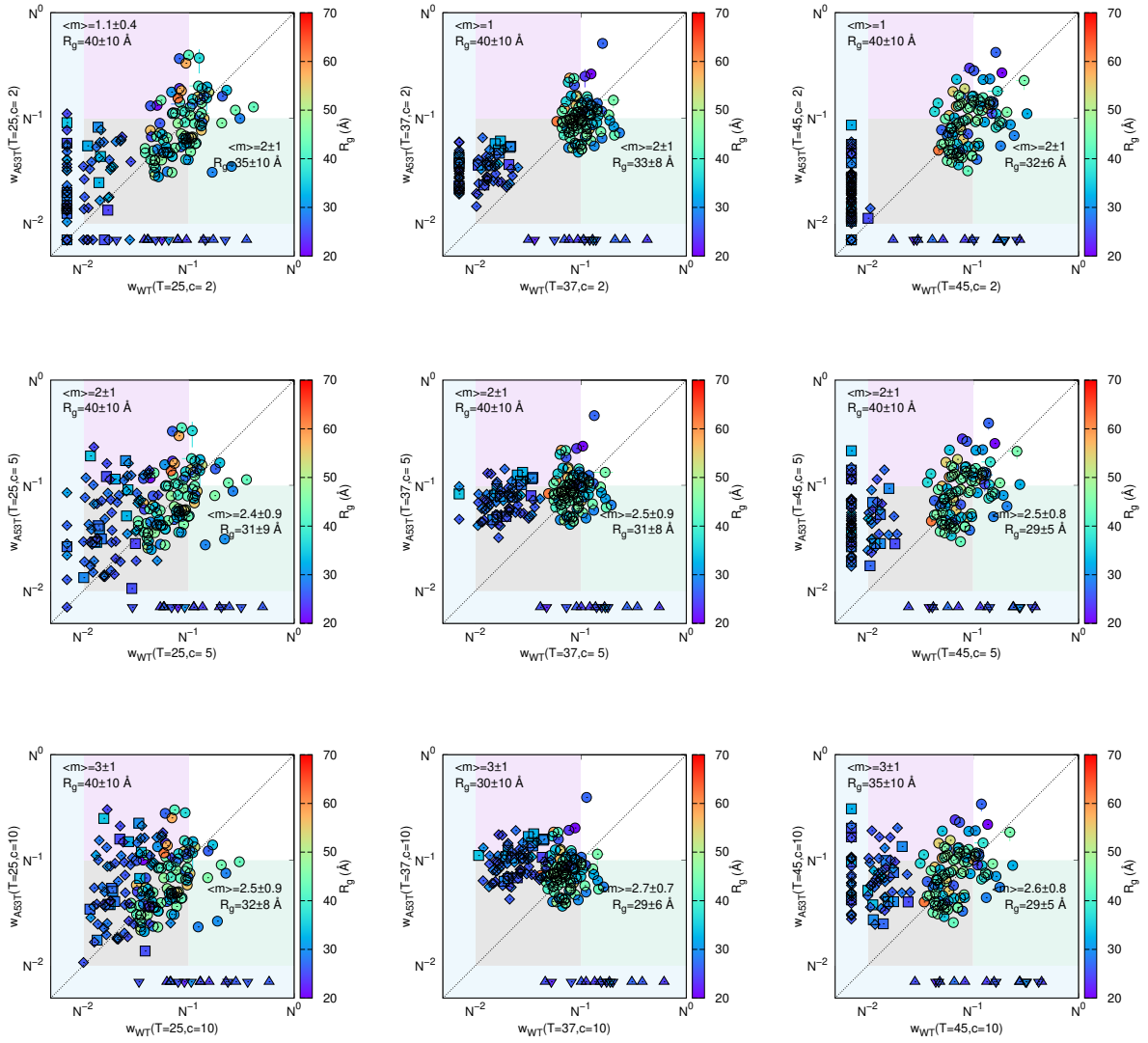


FIGURE 6.24: Correlation map of the weights of the $N = 189$ conformations (Gurry's ensemble) between WT (horizontal axis) and A53T (vertical axis). Each symbol represents the i-conformation and has been color coded on the basis of its R_g, according to the color palette on the right. Symbols are assigned according to Gurry's classification shown in Fig. 4.5. Circles: monomers (A); down-sided triangles: helical-rich trimers (B); up-sided triangles strand-rich trimers (C); squares: helical-rich tetramers (D); diamonds: strand-rich tetramers (E). The purple (light-green) off-diagonal quadrant indicates the group of conformations with w_i simultaneously greater (lower) than N^{-1} for A53T and lower (greater) than N^{-1} for WT. In each off-diagonal quadrant the average values of R_g and $<m>$ of the group of conformations present in the quadrant are reported. Conformations with $w_i < N^{-2}$ for A53T and WT are shown in the middle of the left and bottom cyan strips, respectively. Rows from top to bottom refer to $c = 2$, 5 and 10 g/L. Columns from left to right refer to $T = 25$, 37 and 45°C.

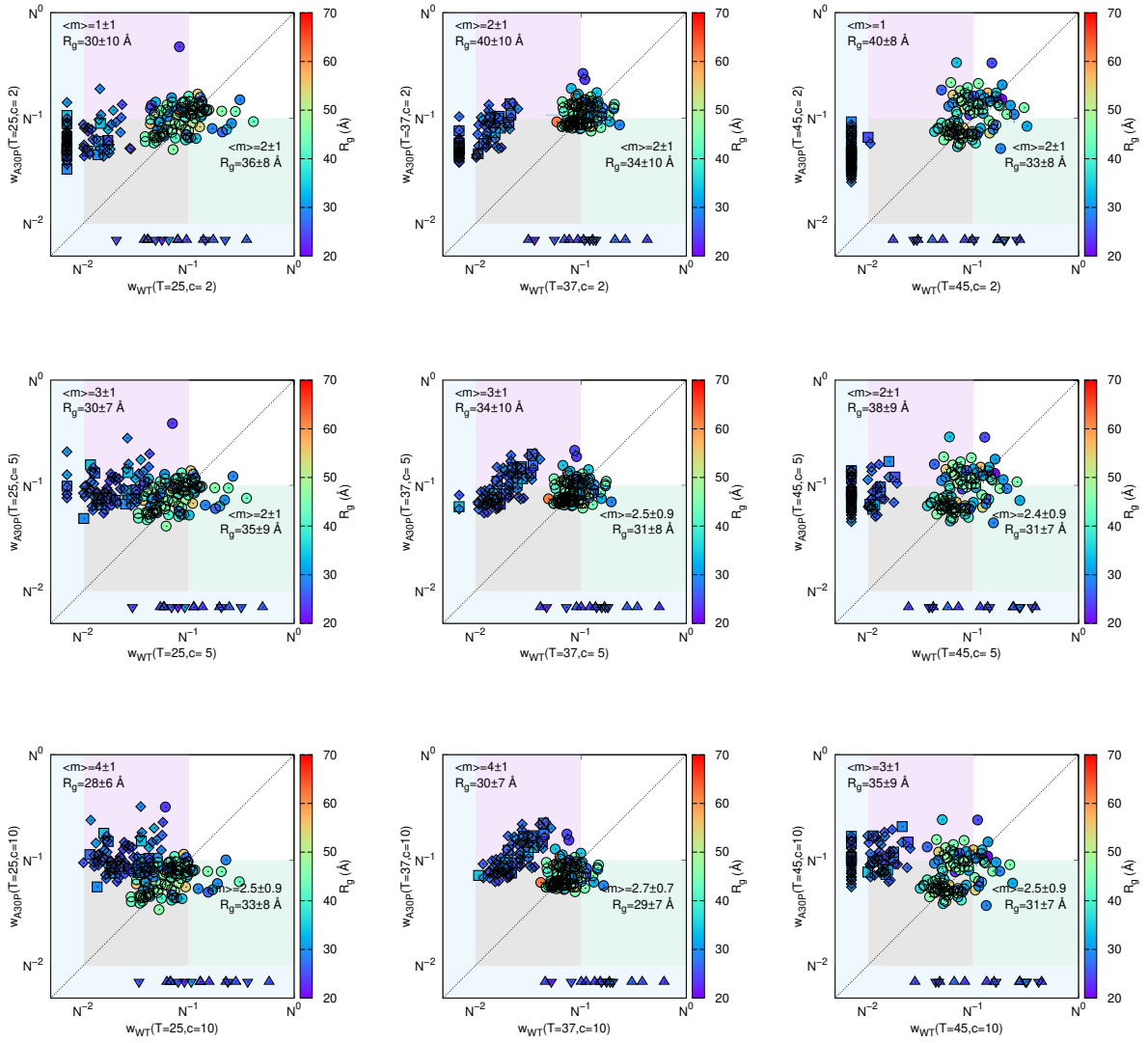


FIGURE 6.25: Correlation map of the weights of the $N = 189$ conformations (Gurry's ensemble) between WT (horizontal axis) and A30P (vertical axis). Each symbol represents the i-conformation and has been color coded on the basis of its R_g, according to the color palette on the right. Symbols are assigned according to Gurry's classification shown in Fig. 4.5. Circles: monomers (A); down-sided triangles: helical-rich trimers (B); up-sided triangles strand-rich trimers (C); squares: helical-rich tetramers (D); diamonds: strand-rich tetramers (E). The purple (light-green) off-diagonal quadrant indicates the group of conformations with w_i simultaneously greater (lower) than N^{-1} for A30P and lower (greater) than N^{-1} for WT. In each off-diagonal quadrant the average values of R_g and $<m>$ of the group of conformations present in the quadrant are reported. Conformations with $w_i < N^{-2}$ for A30P and WT are shown in the middle of the left and bottom cyan strips, respectively. Rows from top to bottom refer to $c = 2$, 5 and 10 g/L. Columns from left to right refer to $T = 25$, 37 and 45°C.

The results shown in these maps show that in most cases the mutants are described by tetrameric conformations, while the WT α -syn is preferentially described by monomeric and trimeric conformations. This is very important result. In fact, the literature reports that the mutants provoke a speed of fibrillation greater than the one of WT. This different aggregation rate could explain the result that we have obtained. The VBWSAS 3.0 method evaluates important tetrameric conformations to describe the SAXS curves of the mutants, while prefers the trimeric conformation in the description of the WT curves. It is also important to note that, from a qualitative analysis, in the description of the mutants there are more strand-rich tetramers (diamonds in the maps), compared to helix tetramers (squares). This result is in agreement with Ref. [17], which describes strand-rich tetramers as the most important conformation involved in the aggregation process.

6.3.6 Thermodynamic parameters

In this section the thermodynamic parameters derived from VBWSAS 3.0 analysis are reported in a tabular form.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
2	1	34±7	39±2	24±7	22±5	−10±6
3	1	3±9	36±2	−20±4	13±5	39.0±0.3
4	1	−40±4	−40±5	−34±4	−24±7	−38±2
5	1	−22±7	−40.0±0.8	−20±10	−38.6±0.8	40±1
6	1	39±3	30±4	−33±3	38±1	−31±3
7	1	39±5	40±10	12±5	40±2	40±1
8	1	40±4	30±10	38±2	17±5	6±5
9	1	28±7	5±4	13±6	24±7	31±4
10	1	−34±2	−34±4	−40±2	−38±1	−38±5
11	1	19±7	40±1	11±7	37±2	34±4
12	1	−25±7	−35±2	−23±5	−17±6	−31±7
13	1	−39±6	−40±3	−38.7±0.8	−39±2	−34±2
14	1	29±3	19±5	−2±7	40±5	34±2
15	1	−37±3	−40.0±0.8	34±7	−38±1	−38±4
16	1	−37±2	−40±5	−36±2	−40±1	−9±9
17	1	−21±7	−19±8	−7±10	−4±4	−26±7
18	1	39±5	27±5	39±1	40±5	40±2
19	1	12±9	15±3	38±5	32±6	10±10
20	1	−14±10	−40±10	−36±3	−34±2	−39±1
21	1	−40±4	−22±4	−39±5	7±5	−34±3
22	1	31±4	38±2	−10±4	39±8	21±8
23	1	−32±9	36±1	−37±2	−39.9±0.3	−27±6
24	1	−40±3	29±4	38±2	−39±1	−38±2
25	1	12±5	31±4	−10±10	9±6	0±6
26	1	14±5	−12±4	−8±7	−10±9	−38.9±0.8
27	1	−35±2	−40.0±0.6	23±5	29±3	39±2
28	1	−37±7	−39±3	−12±5	−32±3	−39.4±0.9
29	1	22±8	12±5	−40±6	39±2	−32±4
30	1	40±1	39±1	28.6±0.5	37±3	40±2
31	1	−14±7	−37±1	−36±2	−26±5	−40±3
32	1	−11±5	−33±4	39±8	39±6	−30±10
33	1	22±5	−1±5	−40±4	29±5	39±2

TABLE 6.1: Dimensionless variations of reference enthalpy between the 1st and the i^{th} conformer derived by VBWSAS 3.0 from SAXS data of WT, G51D, E46K, A53T and A30P α-syn mutants.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
34	1	−34±2	−37±3	−39±2	−38.2±0.7	−37±1
35	1	−33±3	−29±4	−39.2±0.6	−21±5	−39.6±0.4
36	1	−37±6	−40±5	−39±3	−37.7±0.9	−31±7
37	4	1.7±0.7	3±1	8±1	5.3±0.8	−1.1±0.9
38	4	−18.4±0.6	−10.7±0.5	4.3±0.6	−20±1	−22.0±0.3
39	1	40±2	40±1	33±6	40±3	−6±8
40	4	−9±1	−12.7±0.5	−1.3±0.4	−16.3±0.8	−14±1
41	4	9±1	9.8±0.3	8.8±0.8	10.4±0.7	15.7±0.6
42	4	−26.1±0.9	−31±2	−31.0±0.5	−32±1	−18.9±0.6
43	4	2±1	9±1	2±1	1.7±0.7	0.1±0.3
44	1	10±7	29±4	−40±5	−36±2	−40±6
45	4	6.2±0.9	6±2	13.9±0.1	−3.8±0.3	8±1
46	4	−15.1±0.4	−26.0±0.8	−21.9±0.5	−24±1	−32.6±0.2
47	4	13±1	17.8±0.9	5.9±0.9	11.4±0.5	22.3±0.5
48	4	−25.0±0.4	−29±1	−33±1	−17.6±0.9	−33.3±0.7
49	4	−12±1	−17±1	−9.2±0.9	−26.3±0.8	−18.5±0.5
50	4	−1±1	−10.6±0.7	3.9±0.6	−7±1	−14.8±0.9
51	1	39±3	39±1	35±2	39±2	30±4
52	4	−21.6±0.9	−27±1	−19±2	−18.0±0.9	−24.0±0.2
53	4	15.1±0.7	18.5±0.9	21±2	19.0±0.4	22±2
54	4	−24±1	−20.3±0.6	−11.8±0.7	−21±1	5.1±0.5
55	4	4±2	6±2	19.0±0.6	−4.4±0.4	−4.1±0.5
56	4	8.5±0.9	6.1±0.4	3.0±0.5	15±2	−9.5±0.8
57	4	−17±1	−14.3±0.8	−18±1	−20±2	−8.3±0.7
58	1	−35±3	−40±4	−35±3	−39±7	−37±3
59	4	−15.6±0.7	−19±1	−6±1	−11.0±0.7	−27.3±0.5
60	4	−22±1	−18.3±0.9	−26±4	−27±1	−15±2
61	4	−11±1	−9.7±0.8	−2±1	−13±1	−9.2±0.3
62	4	−39.3±0.6	−39.4±0.4	−35±1	−39.0±0.5	−39.0±0.5
63	4	1±2	−4±1	11±3	1.3±0.4	3±1
64	4	6.0±0.6	2.3±0.7	14.7±0.5	0±1	4.0±0.4
65	4	−14.0±0.8	−7±1	−5±1	−19.9±0.8	−19±1

Table 6.2: Continuation of previous Table.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
66	4	−7.6±0.8	−8.2±0.5	−1±1	−17±1	−3.1±0.2
67	4	8±2	2.3±0.6	3.0±0.8	17.1±0.9	−1.3±0.3
68	4	−33±1	−27±2	−22.3±0.4	−37.6±0.5	−25±1
69	4	13±1	9±1	−8.4±0.8	11±1	−2±1
70	4	−4±2	−4±1	0±1	−16.2±0.5	−3.8±0.4
71	4	7±2	8±2	15.9±0.1	7.1±0.6	17.6±0.7
72	4	6±1	13.3±0.5	−1±1	2.7±0.7	−1.0±0.7
73	4	10±2	2±1	18.7±0.3	8.1±0.8	19.1±0.3
74	4	−6.2±0.7	−14.5±0.6	−17±2	−1.1±0.4	−21.0±0.8
75	4	13.0±0.7	11.9±0.5	−3±2	13.7±0.9	−2.5±0.5
76	4	12.9±0.5	9.6±0.5	17.3±0.8	16.4±0.3	8.9±0.4
77	4	−17±2	−28±1	−28±2	−34.9±0.9	−29.3±0.6
78	4	−36.2±0.4	−32.1±0.7	−35.7±0.7	−32±1	−34±1
79	4	−11.2±0.5	−11±1	−17±1	−3.3±0.3	−21±1
80	1	−17±9	−34±8	−40±4	−38±2	−37±9
81	4	−1±1	3.4±0.3	18±1	4±1	10.6±0.5
82	4	−2.6±0.7	−9.5±0.6	−7.1±0.5	5±2	5.0±0.8
83	4	13±1	10.4±0.5	14.1±0.8	7±1	−4.1±0.4
84	4	−30.8±0.7	−39.1±0.5	−38±1	−37±2	−39.5±0.2
85	4	−30±1	−27.6±0.8	−22.3±0.7	−31.1±0.4	−35.0±0.3
86	4	−17±1	−26±1	−24±2	−12.8±0.5	−9.3±0.4
87	4	−8.8±0.8	−16.1±0.5	0±1	−1.8±0.9	−7.7±0.6
88	4	−20.6±0.8	−24±2	−39.8±0.5	−20±1	−20.7±0.7
89	1	36±5	10±10	28±6	36±2	−22±5
90	4	4±1	4.4±0.7	−4±1	−5±2	−0.7±0.8
91	4	−9.7±0.6	−10±1	−24±1	−15±1	−27±1
92	4	16±1	8.6±0.6	25.7±0.7	4±1	5±2
93	4	−1.3±0.8	−4±2	−1±1	−9±1	−15±1
94	4	−24.8±0.7	−22±1	−32.5±0.7	−32.4±0.6	−31.7±0.8
95	4	1±1	0±1	12.8±0.3	11.7±0.9	−4±1
96	4	−17±1	−16.8±0.7	−17.8±0.4	−12.7±0.4	−14.6±0.8
97	4	−2±1	−11.0±0.4	−4±2	−0.3±0.1	−0.2±0.6

Table 6.3: Continuation of previous Table.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
98	4	−22±1	−19±1	−20±1	−22±1	−12.2±0.9
99	4	12±1	3±3	8±2	10.0±0.7	5.5±0.2
100	4	−23±1	−21.4±0.7	−2±2	−31.5±0.7	−14.7±0.8
101	4	16.7±0.8	21.5±0.9	24±1	11±1	14.2±0.3
102	1	−37±6	−22±9	−11±4	0±10	−7±6
103	4	−18.9±0.5	−19±1	−7±1	−22.0±0.4	−10±1
104	4	−7.4±0.8	−10.0±0.5	−2±1	−9±1	−24±1
105	4	9.8±0.7	9±2	5.4±0.7	11.4±0.3	20.2±0.9
106	4	5.7±0.9	1.9±0.6	−8±1	−6.0±0.4	−11.6±0.4
107	4	−2±2	0.9±0.2	−22±1	0.5±0.3	−16.6±0.4
108	4	−15.4±0.9	−18.8±0.6	−1.9±0.6	−20.7±0.8	2.3±0.5
109	4	3±2	1.7±0.7	−1±1	7±1	−0.2±0.5
110	1	−39±1	20±3	40±1	−40±3	−29±4
111	4	−18.6±0.6	−19±2	−28.7±0.8	−22±1	−6.9±0.3
112	4	−37.8±0.9	−35.3±0.5	−38.8±0.8	−32.9±0.3	−27±1
113	4	−0.6±0.8	1±1	2.0±0.7	−10±1	14.8±0.5
114	4	−3±1	−4.8±0.7	5±2	−1.7±0.6	4.8±0.2
115	4	2.0±0.8	6.8±0.5	3.6±0.9	−0.1±0.3	7±1
116	4	−24±1	−26.6±0.6	−24.9±0.2	−24.3±0.2	−28±1
117	4	−12.7±0.3	−17.8±0.8	−34±1	−15.0±0.3	−15.9±0.3
118	4	21±2	22.7±0.8	12.3±0.9	28.9±0.8	24.4±0.7
119	4	−25±1	−32±1	−28.5±0.6	−20.4±0.5	−11.9±0.4
120	4	−19.4±0.9	−16.4±0.7	−11.4±0.8	−12±1	−19.9±0.6
121	3	18±3	38±1	36±1	34±2	19±1
122	3	20±2	27±3	16.9±0.9	14±3	34.2±0.6
123	3	−31±2	−28±1	−38±4	−38±2	−23±3
124	3	4.1±0.9	23±1	18.9±0.7	−2±1	5±1
125	3	−35±1	−39.8±0.3	−18±1	−33±2	−8±2
126	3	−32±2	−21±2	−38.8±0.4	−33±2	−35±2
127	3	−40.0±0.7	−37±1	−37±1	−38±1	−34±1
128	3	−3±1	21±2	2±1	8±2	7±2
129	1	−36±1	−39±3	−39±1	−29±5	−35±2

Table 6.4: Continuation of previous Table.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
130	3	−24±2	−2±2	−9±2	−24±3	−33±1
131	3	−40±2	−34±2	−38.0±0.9	−18±4	−32±1
132	3	12.9±0.3	10.2±0.7	20±2	40.0±0.6	36±1
133	3	−8±2	−5±2	−39.3±0.7	−30±3	−34±1
134	3	−21.2±0.8	−5±2	−5±2	−40±3	−14±2
135	3	7.0±0.9	−5±2	29±2	20±1	1±3
136	1	−40±3	−31±4	−26±4	−39±10	−26±1
137	3	31.7±0.7	39±1	8±2	40±2	24.5±0.9
138	1	39.2±0.9	40±1	36±8	39±1	−40±1
139	1	−38±4	−29±10	36±7	−40±7	39±4
140	1	38.8±0.5	34±5	−34±7	−38±3	−7±9
141	1	40±8	40±2	−10±10	34±3	−13±5
142	1	−40±2	−17±2	−37±2	−32±6	−40.0±0.9
143	1	24±7	15±8	−40±2	34±7	40±2
144	1	−34±4	−29±5	−1±9	20±10	−35±2
145	1	−35±2	−9±4	−40±2	−40±6	−31±5
146	1	−38±1	−31±6	−26±6	−34±6	−34±8
147	1	4±5	1±8	−37±10	11±7	−39±2
148	1	−11±2	−38.4±0.7	−39±8	−33±10	40±2
149	1	40±3	37±2	29.8±0.8	40±3	−24±3
150	1	−39.9±0.4	−40±5	−40±3	−38±2	−28±5
151	1	−18±6	23±5	40±3	37±4	35±4
152	1	27±3	33±4	39±2	35±5	38±2
153	1	−39±2	−39±7	−10±5	−40±10	−34±4
154	1	35±4	21±2	2±8	34±5	5±3
155	1	29±4	37±2	35±4	0±3	14±6
156	1	−5±4	4±4	−24±2	11±2	39±3
157	1	10±10	−13±1	40±5	20±10	32±3
158	1	40±2	39.9±0.4	40±5	40±3	40±1
159	1	34±8	21±7	−34±9	3±3	33±3
160	1	−9±5	−40±1	−38±3	−40±1	−19±9
161	1	−40.0±0.7	−34±3	19±1	−34±3	−39±5

Table 6.5: Continuation of previous Table.

i	m_i	$\Delta H^0_{i1}/(RT_0)$				
		wt	g51d	e46k	a53t	a30p
162	1	−29±7	−39±3	−39.8±0.9	−38±2	−38±1
163	1	40±1	28±5	40±1	39.8±0.5	34±7
164	1	38±4	40±5	12±2	40±10	25±6
165	1	−39±3	−39±3	32±4	−40±2	23±6
166	1	−32±4	−39±2	−3±6	−40±4	10±3
167	1	40±3	20±10	40±2	35±2	39.1±0.6
168	1	−39±9	−22±6	−27±6	−8±4	−18±4
169	1	37±1	40±2	35±7	31±6	38±2
170	1	−35±2	−40±3	−27±5	−33±3	−39±4
171	1	−40±4	−33±6	40±2	−38.6±0.5	39.2±0.4
172	1	−40±3	−13±4	−32±3	−40±3	−40±8
173	1	26±7	34±4	12±8	−39±5	−40±9
174	1	40±1	40±5	40±6	34±3	39±9
175	1	13±5	21±4	38±3	13±6	39±2
176	1	26±8	40±4	17±5	38±2	40±4
177	1	2±8	24±5	−20±8	40±1	39±3
178	1	38±5	−2±6	34±3	39±2	5±2
179	1	−30±10	−29±6	−40±10	−40±2	−39±1
180	1	33±6	39±2	1±2	40±2	40±1
181	1	14±7	9±3	40±2	35±3	40±3
182	1	−27±7	1±6	−39.8±0.1	−39.5±0.2	−36±3
183	1	−12±5	−24±5	31±4	−32±3	40±1
184	1	40±4	37±3	−12±9	35±8	−1±4
185	1	10±10	19±6	−20±9	37±2	40±2
186	1	−36±3	4±7	−37±3	−37±1	−28±3
187	1	6±6	−39±9	−15±4	−40±5	−40±3
188	1	−40±1	−34±2	−32±3	−38±3	−40±3
189	1	20±4	38±3	36±4	33±3	35±2

TABLE 6.6: Continuation of previous Table.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
2	1	33±7	39±1	24±7	23±4	−9±6
3	1	3±9	36±2	−20±3	14±5	39.3±0.6
4	1	−40±4	−38±5	−34±4	−22±6	−36±1
5	1	−23±7	−39±1	−20±10	−36.7±0.6	39±1
6	1	37±4	31±4	−33±2	37±2	−30±3
7	1	37±5	40±10	11±4	40±2	40±1
8	1	39±4	30±10	37±2	21±6	7±5
9	1	28±8	6±4	12±6	25±7	32±4
10	1	−35±2	−31±4	−39±2	−36±2	−37±5
11	1	18±7	39.8±0.9	10±7	37±2	34±4
12	1	−25±7	−32±2	−23±4	−13±5	−30±7
13	1	−39±6	−38±3	−38±1	−39±2	−33±2
14	1	28±3	20±5	−4±6	40±6	35±2
15	1	−37±4	−38±1	33±7	−36±2	−36±3
16	1	−37±3	−38±5	−35±2	−37±1	−8±9
17	1	−22±7	−17±8	−8±9	−3±5	−25±6
18	1	38±5	27±5	39±2	40±5	40±2
19	1	12±8	16±3	36±5	33±5	8±10
20	1	−10±10	−40±10	−35±4	−32±2	−38±1
21	1	−39±3	−20±4	−39±5	8±5	−33±3
22	1	29±3	38±2	−11±3	40±9	22±8
23	1	−32±8	36.4±0.9	−38±2	−39.2±0.7	−26±6
24	1	−39±3	30±3	36±2	−36±1	−36±2
25	1	11±5	32±3	−13±10	9±6	1±6
26	1	13±5	−11±4	−8±8	−7±9	−34±1
27	1	−34±3	−38±1	22±5	31±3	39±2
28	1	−36±7	−38±3	−12±6	−30±3	−38±1
29	1	20±7	13±5	−40±7	38±2	−30±4
30	1	38±1	39±1	33±2	40±3	40±2
31	1	−14±6	−34.9±0.7	−36±2	−25±4	−39±4
32	1	−11±4	−31±4	38±8	40±6	−30±10
33	1	20±4	1±5	−40±5	30±4	40±2

TABLE 6.7: Dimensionless variations of reference entropy between the 1st and the i^{th} conformer derived by VBWSAS 3.0 from SAXS data of WT, G51D, E46K, A53T and A30P α-syn mutants.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
34	1	−35±2	−34±3	−39±2	−35±1	−35±1
35	1	−34±3	−28±4	−39±1	−17±5	−38.1±0.8
36	1	−38±5	−38±5	−38±3	−35±1	−30±7
37	4	10.1±0.8	14±1	14±1	16.0±0.7	9.9±0.8
38	4	−9.6±0.4	0.6±0.6	11.2±0.5	−9±1	−11.1±0.5
39	1	39±2	39.9±0.8	32±6	39±3	−5±8
40	4	0±1	−1.2±0.6	5.3±0.4	−5.0±0.6	−2.7±0.9
41	4	17±1	20.8±0.2	16±1	21.2±0.6	26.5±0.5
42	4	−17±1	−20±1	−24.0±0.7	−20±1	−7.8±0.5
43	4	11±1	20.0±0.8	8±1	13.2±0.7	11.3±0.3
44	1	9±6	30±4	−40±6	−33±2	−38±6
45	4	15.1±0.9	17±1	20.6±0.1	7.3±0.4	19±1
46	4	−6.5±0.4	−14.7±0.5	−14.8±0.4	−12±2	−21.4±0.2
47	4	22±1	28.5±0.9	12±1	21.8±0.7	33.0±0.4
48	4	−15.7±0.4	−17±1	−26±1	−6±1	−22.0±0.9
49	4	−4±2	−5±1	−2.7±0.8	−15.3±0.9	−7.5±0.5
50	4	8±1	0.3±0.6	10.9±0.8	4±1	−4.0±0.8
51	1	38±3	39±1	33±3	38±2	31±4
52	4	−13±1	−15±1	−12±2	−7±1	−13.0±0.4
53	4	23.9±0.6	29.2±0.8	28±2	29.4±0.5	32±2
54	4	−15±1	−9.2±0.5	−4.8±0.7	−10±1	16.2±0.4
55	4	12±2	18±2	25.5±0.6	6.4±0.5	6.8±0.5
56	4	17±1	16.8±0.5	9.4±0.5	25±2	1.5±0.7
57	4	−8±1	−2.8±0.7	−11±1	−8±2	2.6±0.6
58	1	−32±2	−37±4	−35±2	−37±8	−35±3
59	4	−7.1±0.8	−7±1	1±1	0.6±0.5	−16.3±0.5
60	4	−13±1	−7.3±0.9	−19±4	−15±1	−4±2
61	4	−2±2	1.3±0.5	5±1	−2.0±0.9	1.7±0.4
62	4	−30.0±0.7	−27.7±0.3	−28±1	−27.4±0.5	−27.6±0.3
63	4	9±2	7±1	17±3	11.8±0.4	13.9±0.8
64	4	14.9±0.8	13.4±0.6	21.0±0.6	10±1	15.1±0.3
65	4	−5.4±0.6	4±1	1±1	−8.6±0.6	−7.9±0.9

Table 6.8: Continuation of previous Table.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
66	4	1.4±0.7	3.0±0.4	5.6±0.5	−5±1	7.9±0.4
67	4	16±1	13.0±0.2	9.8±0.5	27.6±0.8	9.7±0.2
68	4	−23.5±0.8	−16±2	−15.5±0.5	−25.9±0.6	−13±1
69	4	22±2	19±1	−1.7±0.8	21.8±0.8	9.3±0.8
70	4	5±2	7±2	7.2±0.9	−4.7±0.5	7.4±0.3
71	4	16±2	19±2	22.5±0.3	18.0±0.4	28.4±0.6
72	4	14±1	24.1±0.3	6±1	13.7±0.9	10.2±0.6
73	4	18±2	13±2	25.2±0.4	18.6±0.4	29.8±0.4
74	4	2.9±0.7	−3.3±0.7	−10±2	9.9±0.3	−9.9±0.9
75	4	21.4±0.8	22.7±0.4	4±2	24±1	9.0±0.3
76	4	21.5±0.5	20.29±0.07	24±1	27.0±0.2	19.9±0.4
77	4	−8±2	−17±1	−21±2	−24±1	−18.1±0.3
78	4	−27.1±0.4	−20.6±0.5	−28.5±0.6	−20±1	−22±1
79	4	−2.6±0.5	0±1	−11±1	7.7±0.1	−10±1
80	1	−20±10	−32±8	−39±4	−36±2	−35±9
81	4	8.0±0.9	14.0±0.4	24.7±0.9	16±1	21.5±0.4
82	4	6.8±0.5	1.3±0.6	−0.4±0.4	15±2	16±1
83	4	22±1	20.9±0.5	20.5±0.6	17.8±0.7	6.8±0.5
84	4	−21.7±0.8	−27.2±0.3	−30±1	−25±2	−27.9±0.2
85	4	−21±1	−16.1±0.8	−15.3±0.5	−19.8±0.8	−23.7±0.5
86	4	−8±1	−15±2	−17±2	−1.9±0.3	1.6±0.3
87	4	0.3±0.9	−4.9±0.6	7.1±0.9	8.4±0.6	3.2±0.9
88	4	−12.0±0.6	−12±2	−32.5±0.5	−9±1	−9.7±0.6
89	1	36±5	10±10	26±6	36±2	−21±5
90	4	13±1	15.2±0.7	4±1	5±2	10.2±0.9
91	4	−0.8±0.7	0.2±0.7	−18±1	−4±1	−16.4±0.9
92	4	24.7±0.8	19.5±0.5	31.9±0.6	14.4±0.7	16±2
93	4	7.3±0.8	7±1	5.6±0.6	2±2	−4±1
94	4	−16.2±0.8	−10±1	−25.4±0.8	−20.5±0.8	−20.5±0.8
95	4	10±1	11.3±0.8	19.4±0.3	22.0±0.8	6.5±0.9
96	4	−8±1	−6±1	−10.7±0.3	−1.8±0.1	−4±1
97	4	7±1	0.1±0.2	3±2	11.2±0.6	11.0±0.5

Table 6.9: Continuation of previous Table.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
98	4	−13.2±0.9	−7±1	−13±1	−10±1	−1.5±0.9
99	4	20±1	14±2	15±1	20.8±0.6	16.7±0.2
100	4	−15±1	−10.1±0.6	4±2	−20.2±0.7	−3.8±0.6
101	4	25.4±0.5	32±1	30±1	22±1	25.0±0.2
102	1	−35±6	−20±8	−10±3	0±10	−7±6
103	4	−9.5±0.6	−8±2	0±1	−10.6±0.6	1±1
104	4	1.4±0.5	1.4±0.4	5±1	2.8±0.6	−13±1
105	4	18.4±0.7	20±1	11.8±0.8	22.2±0.7	31.0±0.9
106	4	14.2±0.9	12.6±0.6	−1±2	5.4±0.5	−0.4±0.4
107	4	7±2	11.7±0.7	−15±1	10.9±0.4	−5.8±0.3
108	4	−6.6±0.9	−7.7±0.5	4.9±0.5	−9.8±0.9	13.4±0.5
109	4	12±1	13.1±0.6	6±1	18±1	10.7±0.6
110	1	−39±1	21±2	38±1	−37±3	−27±4
111	4	−9.8±0.8	−8±2	−21.7±0.8	−10.5±0.9	4.0±0.4
112	4	−28.5±0.6	−23.7±0.5	−31.5±0.5	−21.2±0.3	−16.2±0.7
113	4	8.5±0.7	12±1	8.6±0.7	2±1	25.6±0.6
114	4	6±2	7.1±0.7	12±2	9.8±0.7	15.6±0.4
115	4	11.2±0.8	17.5±0.2	10.2±0.7	10.4±0.3	18±1
116	4	−16±1	−15.1±0.3	−18.0±0.2	−12.9±0.2	−17.0±0.9
117	4	−4.2±0.5	−6.3±0.7	−27±2	−4.2±0.4	−5.0±0.5
118	4	29±1	33.6±0.6	18.8±0.7	39.2±0.7	35.0±0.8
119	4	−16±1	−20.4±0.8	−21.4±0.5	−9.6±0.3	−1.1±0.5
120	4	−11±1	−5.6±0.5	−4.8±0.7	−1±1	−8.9±0.6
121	3	27±3	37±1	34±1	32±2	23±1
122	3	30±2	27±3	15.4±0.8	12±3	37.5±0.9
123	3	−22±1	−28±1	−39±4	−38±2	−19±2
124	3	13±1	22.6±0.9	17±1	−3±2	9±2
125	3	−26±1	−39.6±0.9	−20±1	−33±1	−4±2
126	3	−22±2	−21±2	−39.7±0.5	−33±2	−31±2
127	3	−30.7±0.9	−37.0±0.9	−38.2±0.6	−38±1	−29±1
128	3	5±2	20±2	−0.2±0.9	6±2	11±2
129	1	−36±1	−36±2	−37±1	−26±5	−34±2

TABLE 6.10: Continuation of previous Table.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
130	3	−14±2	−3±1	−10±1	−24±2	−28±1
131	3	−31±2	−34±2	−39.1±0.3	−19±4	−28±1
132	3	20.6±0.5	9±1	17±3	37.6±0.6	39.5±0.8
133	3	1±2	−5±2	−40±1	−31±3	−30±1
134	3	−12.3±0.6	−5±1	−6±1	−40±3	−10±1
135	3	15±1	−4±2	27±2	18±1	5±3
136	1	−39±2	−29±3	−25±4	−36±9	−25±1
137	3	40±1	38±1	6±2	38±2	28.0±0.8
138	1	38.3±0.8	40.0±0.9	35±8	39±1	−38±1
139	1	−36±3	−27±10	36±6	−37±7	39±4
140	1	37±1	34±5	−34±8	−35±3	−6±9
141	1	39±7	40±2	−10±10	34±3	−11±4
142	1	−39±2	−15±2	−37±2	−29±5	−39±1
143	1	24±7	16±7	−40±2	35±6	40±2
144	1	−34±4	−27±5	−1±8	20±10	−34±2
145	1	−34±3	−8±5	−39±2	−37±6	−30±6
146	1	−37±1	−29±6	−26±5	−31±6	−32±9
147	1	5±6	2±7	−38±10	12±6	−37±2
148	1	−11±2	−36±1	−39±8	−30±10	40±2
149	1	39±3	37±1	28±1	40±2	−23±4
150	1	−39.3±0.7	−38±5	−39±2	−37±2	−27±4
151	1	−17±7	24±5	39±3	36±4	35±4
152	1	26±3	33±4	38±3	35±5	38±2
153	1	−38±2	−37±7	−9±4	−30±10	−33±5
154	1	34±3	22±2	1±6	33±5	7±4
155	1	28±5	36±2	32±3	−1±3	15±6
156	1	−5±4	5±4	−23±2	12±2	39±3
157	1	10±10	−12±1	40±4	20±10	32±4
158	1	40±2	39.9±0.6	39±5	40±2	40±1
159	1	33±8	23±6	−34±9	5±4	33±3
160	1	−9±5	−39±2	−38±3	−39±2	−17±9
161	1	−39.5±0.8	−31±2	19±2	−32±2	−37±4

TABLE 6.11: Continuation of previous Table.

i	m_i	$\Delta S^0_{i1}/R$				
		wt	g51d	e46k	a53t	a30p
162	1	−29±7	−36±3	−40±2	−36±2	−37±1
163	1	38±2	31±4	39±1	39.9±0.9	34±7
164	1	36±4	40±5	12±1	39±9	26±6
165	1	−36±3	−37±2	30±5	−38±2	24±6
166	1	−33±4	−36±2	−3±6	−38±4	11±4
167	1	38±3	20±10	38±2	34±3	38.8±0.5
168	1	−39±9	−21±5	−26±7	−8±3	−17±4
169	1	35±2	39±1	33±6	32±5	40±2
170	1	−35±2	−38±3	−27±5	−31±3	−37±4
171	1	−38±4	−30±5	38±2	−35.8±0.9	40±1
172	1	−40±2	−11±4	−32±3	−39±3	−38±8
173	1	25±8	34±4	13±8	−39±5	−40±10
174	1	38±2	40±5	39±6	35±2	39±9
175	1	10±4	22±4	37±4	13±6	39±1
176	1	24±7	40±4	15±4	38±2	40±4
177	1	1±7	26±4	−20±7	39±1	40±3
178	1	37±5	−1±5	33±3	40±1	5±2
179	1	−28±9	−27±6	−40±10	−38±2	−38±2
180	1	31±6	39±3	0±3	40±1	40±1
181	1	12±7	11±3	38±2	36±2	40±2
182	1	−27±7	1±5	−40±1	−35±2	−34±2
183	1	−12±5	−23±5	31±3	−30±4	40±1
184	1	38±4	38±4	−13±9	35±7	−1±4
185	1	10±10	20±6	−19±8	37±1	40±2
186	1	−34±4	5±6	−37±3	−35±1	−27±3
187	1	5±6	−37±9	−16±3	−37±4	−39±3
188	1	−38±2	−31±2	−31±3	−36±3	−38±3
189	1	19±5	38±3	34±3	34±2	35±2

TABLE 6.12: Continuation of previous Table.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
2	1	−200±400	−100±200	−600±200	−200±500	300±500
3	1	−100±400	−600±500	400±300	100±200	−200±200
4	1	600±500	500±500	600±300	−600±500	400±500
5	1	400±200	600±200	−600±500	600±500	300±400
6	1	100±400	300±500	500±90	−300±300	600±300
7	1	−200±300	−200±600	400±300	−400±100	−600±300
8	1	−500±300	−600±300	−500±300	−600±400	300±400
9	1	400±500	−200±400	−300±400	400±400	−600±400
10	1	500±200	−100±400	−400±500	−500±400	600±400
11	1	500±400	−100±300	−500±500	−100±300	−500±200
12	1	500±300	−100±400	300±400	−600±400	−200±400
13	1	500±400	600±500	−400±400	600±500	500±200
14	1	−600±500	−600±200	500±400	−600±500	−400±200
15	1	600±300	500±200	−600±400	400±400	200±200
16	1	500±400	500±30	−300±500	600±400	−400±400
17	1	500±300	400±200	600±500	−200±400	200±300
18	1	−300±400	200±200	−600±200	400±400	−600±200
19	1	−600±500	−600±100	−600±500	−500±500	500±500
20	1	−300±400	500±500	500±400	300±400	600±500
21	1	300±500	200±500	600±400	−100±500	300±300
22	1	−500±400	−300±200	600±500	−300±200	−500±400
23	1	300±300	−300±100	600±400	600±300	600±40
24	1	600±400	400±300	400±500	300±100	−200±500
25	1	−300±300	−500±200	480±60	−600±300	−500±500
26	1	−400±500	400±300	−400±400	600±500	−600±500
27	1	200±300	500±400	600±500	−300±500	−600±500
28	1	600±600	400±400	500±400	−500±400	500±200
29	1	0±400	−600±300	500±500	500±400	400±400
30	1	200±500	−600±300	−600±70	−600±400	−400±300
31	1	100±400	600±400	500±200	400±200	600±400
32	1	600±400	500±300	600±500	−600±70	400±300
33	1	500±300	−400±400	300±300	−600±200	−600±500

TABLE 6.13: Dimensionless variations of heat capacity at constant pressure between the 1st and the i^{th} conformer derived by VBWSAS 3.0 from SAXS data of WT, G51D, E46K, A53T and A30P α-syn mutants.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
34	1	600±500	200±400	500±400	100±300	500±200
35	1	600±300	600±300	600±200	−600±500	600±100
36	1	500±300	600±500	−300±300	−300±300	300±400
37	4	−100±200	−300±400	−600±300	−500±200	−270±100
38	4	500±400	100±200	−500±200	0±200	600±100
39	1	200±500	−600±400	−500±500	−600±400	100±100
40	4	−100±400	−100±300	−100±300	300±100	400±300
41	4	−570±30	−300±200	−600±200	−600±200	−600±80
42	4	600±300	490±60	500±100	500±100	300±200
43	4	−200±100	−500±200	−200±100	−600±300	−500±200
44	1	500±400	−600±300	200±400	600±300	300±300
45	4	−500±300	−400±300	−540±80	−400±100	−310±100
46	4	300±200	400±200	430±70	0±300	590±70
47	4	−580±70	−589±10	−200±200	−400±100	−600±200
48	4	100±100	600±200	530±50	0±400	580±80
49	4	500±400	200±200	300±500	600±100	300±100
50	4	−200±100	400±200	−400±300	−200±200	400±300
51	1	−300±400	−300±400	−500±500	−100±300	−500±400
52	4	400±200	590±70	0±300	600±300	500±200
53	4	−600±100	−400±200	−590±20	−580±80	−590±30
54	4	600±200	600±200	−200±200	600±100	−500±200
55	4	−500±500	−200±200	−600±40	0±300	−100±100
56	4	−450±20	0±300	−400±200	−600±200	100±200
57	4	600±300	0±200	−100±200	100±200	100±100
58	1	400±400	200±400	470±60	300±400	−200±400
59	4	300±100	600±300	100±300	−500±400	580±40
60	4	400±200	600±200	540±60	400±100	200±100
61	4	−100±300	100±300	−400±400	0±300	0±300
62	4	600±20	570±40	500±50	410±70	600±20
63	4	−300±300	−200±300	−600±300	100±300	−300±200
64	4	−500±400	−300±200	−200±100	−300±200	−500±200
65	4	300±500	−200±300	200±200	300±300	400±200

TABLE 6.14: Continuation of previous Table.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
66	4	0±100	0±200	−300±400	−100±300	−300±300
67	4	−500±100	0±400	−300±200	−520±40	−300±400
68	4	400±200	500±200	600±200	600±200	300±200
69	4	−520±50	−100±300	0±400	−500±100	−300±200
70	4	−400±400	100±400	−500±300	0±200	−300±500
71	4	−600±100	−400±200	−520±100	−600±100	−600±70
72	4	−300±200	−500±100	−300±200	−100±200	−400±200
73	4	−520±30	−200±100	−600±50	100±300	−570±30
74	4	−200±300	300±200	200±300	−100±200	300±100
75	4	−560±40	−330±90	−400±100	−550±70	−500±400
76	4	−600±100	−100±300	−500±100	−600±200	−600±400
77	4	−100±300	500±100	600±100	600±200	600±100
78	4	600±200	600±100	590±40	490±90	560±90
79	4	400±400	200±100	600±300	−200±400	530±50
80	1	400±200	500±300	300±400	−400±400	300±200
81	4	−500±300	100±400	−590±70	−600±200	−600±90
82	4	−600±300	300±300	0±300	−300±200	−400±100
83	4	−590±80	−100±200	−520±90	−200±200	100±200
84	4	590±60	500±50	400±80	600±70	590±20
85	4	510±70	400±200	100±200	600±300	580±100
86	4	300±400	600±200	300±200	300±200	200±200
87	4	−300±200	400±200	−600±80	500±400	100±200
88	4	600±300	300±200	600±100	600±200	400±100
89	1	−500±400	200±500	200±300	−200±400	100±300
90	4	−600±400	100±200	−500±200	300±300	−100±200
91	4	−100±300	500±300	300±100	600±400	530±90
92	4	−550±30	−300±100	−600±20	400±400	−400±100
93	4	−200±300	100±200	−230±90	600±400	500±300
94	4	600±300	0±300	500±100	300±100	500±200
95	4	−200±300	300±400	−600±90	−300±100	100±300
96	4	400±200	400±300	200±200	0±300	300±400
97	4	−300±400	200±300	200±400	−600±400	−300±300

TABLE 6.15: Continuation of previous Table.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
98	4	300±400	0±200	0±300	−200±300	400±300
99	4	−500±100	−100±100	−600±100	−300±200	−600±100
100	4	600±200	500±80	100±300	490±70	200±200
101	4	−600±90	−590±40	−600±100	−530±90	−600±100
102	1	−600±500	400±500	−100±300	600±500	500±300
103	4	100±200	600±100	−200±400	100±200	100±300
104	4	−200±400	−100±300	−200±300	−500±400	600±100
105	4	−500±200	−400±200	−200±200	−600±200	−600±100
106	4	−400±300	100±400	−300±300	−500±600	−100±300
107	4	−300±100	0±400	400±100	300±300	470±50
108	4	300±200	520±30	−300±200	500±400	−500±100
109	4	−500±300	−500±200	−400±300	−400±100	−200±300
110	1	490±90	300±300	200±300	500±500	−200±300
111	4	300±200	400±200	500±200	500±400	0±300
112	4	600±200	600±100	600±200	590±90	600±100
113	4	−600±400	−400±200	−400±200	−400±400	−540±40
114	4	−100±200	−500±300	−600±100	−600±400	−400±200
115	4	−500±100	100±300	−400±100	200±200	−400±200
116	4	400±200	400±100	420±70	400±300	530±50
117	4	300±200	0±400	570±60	300±300	400±100
118	4	−590±40	−580±90	−590±70	−600±30	−600±40
119	4	500±300	300±100	400±100	600±100	320±90
120	4	500±300	600±300	200±200	500±100	400±100
121	3	−300±300	−590±30	−600±100	−560±30	−600±100
122	3	−440±60	−580±90	−600±200	−300±200	−590±40
123	3	600±50	593±9	600±400	590±80	500±100
124	3	−400±400	−600±300	−100±300	−100±500	−600±300
125	3	300±100	560±90	600±300	500±200	200±400
126	3	600±200	400±100	400±100	400±200	600±200
127	3	550±50	300±100	600±200	500±100	600±300
128	3	500±300	−600±200	600±500	300±300	−400±500
129	1	100±300	−300±200	500±400	600±300	500±500

TABLE 6.16: Continuation of previous Table.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
130	3	400±100	−100±200	−300±400	0±300	560±10
131	3	580±90	600±300	570±70	100±200	600±200
132	3	500±500	100±200	0±300	−510±60	−600±50
133	3	600±500	200±200	300±100	300±100	490±50
134	3	520±60	300±300	−600±400	540±20	200±200
135	3	−200±300	−600±300	−600±100	−600±200	0±400
136	1	300±500	100±500	−600±600	−300±500	−200±400
137	3	−500±400	−500±100	−500±200	−560±90	−600±100
138	1	400±400	400±400	300±100	−400±400	600±400
139	1	0±300	100±500	−600±200	−300±500	−600±300
140	1	−500±500	−470±70	500±300	300±300	−500±500
141	1	−600±500	−600±300	−300±300	−200±400	−200±200
142	1	400±400	200±400	400±400	−600±600	600±300
143	1	−600±500	500±500	300±500	−300±100	−400±300
144	1	600±400	300±600	−200±600	−500±500	500±300
145	1	100±400	0±300	200±200	300±300	500±100
146	1	400±400	−300±400	600±500	−500±500	600±500
147	1	−600±400	400±300	500±400	0±400	300±100
148	1	−200±400	600±40	500±400	500±400	−580±80
149	1	−600±500	−300±300	300±300	300±400	600±200
150	1	500±300	600±500	300±200	0±400	600±500
151	1	−400±400	−300±400	−500±400	500±500	−540±20
152	1	400±400	500±500	−600±500	−600±500	−100±400
153	1	−100±300	600±200	−500±300	−300±500	400±400
154	1	−200±300	−600±300	500±500	−200±500	−100±500
155	1	0±400	−100±400	600±500	300±500	−600±500
156	1	−100±400	400±500	−500±500	600±200	−500±400
157	1	−600±500	200±300	−400±400	−600±500	−430±90
158	1	−400±300	−600±300	−600±500	−600±100	−600±400
159	1	−100±300	−600±300	600±500	−600±400	−500±400
160	1	300±600	600±100	500±200	600±300	−500±500
161	1	300±300	500±300	−100±400	400±400	570±80

TABLE 6.17: Continuation of previous Table.

i	m_i	$\Delta C_{p_{i1}}/R$				
		wt	g51d	e46k	a53t	a30p
162	1	600±500	−100±300	500±500	−200±300	200±100
163	1	−100±200	−600±300	−300±100	−300±500	−500±400
164	1	300±500	−500±400	−300±500	−500±500	−600±300
165	1	−600±500	−200±300	−300±200	−500±500	−500±100
166	1	200±200	−100±300	300±300	500±300	−600±500
167	1	−200±400	−300±500	−300±400	400±400	−500±200
168	1	600±400	600±500	400±200	500±600	200±300
169	1	100±400	500±500	600±600	−600±200	−300±300
170	1	600±200	600±500	100±300	400±400	300±400
171	1	−400±400	500±400	400±300	200±200	−300±300
172	1	400±400	−400±400	600±500	400±200	400±500
173	1	−400±400	−500±400	100±300	500±400	300±500
174	1	−600±300	−600±500	−600±600	−500±500	−600±400
175	1	500±500	−600±400	−600±400	400±300	−200±300
176	1	100±500	0±400	500±400	−400±200	−600±500
177	1	−100±500	−600±500	200±500	400±600	−500±400
178	1	−500±400	−200±400	−200±300	−300±400	−200±300
179	1	−100±300	400±300	200±200	300±300	600±100
180	1	100±400	400±400	−600±500	−100±500	−600±500
181	1	200±200	−600±400	−300±400	0±200	−400±300
182	1	400±100	600±500	400±400	−200±500	−200±400
183	1	400±400	600±200	−400±300	600±100	−300±200
184	1	−500±500	−300±500	500±400	100±300	500±500
185	1	0±300	−600±500	100±400	0±300	−600±200
186	1	−500±300	300±300	500±200	400±100	0±300
187	1	−300±400	300±500	100±200	500±300	600±300
188	1	−300±400	100±300	600±100	−200±300	500±200
189	1	−500±200	−500±200	500±500	100±300	400±400

TABLE 6.18: Continuation of previous Table.

Chapter 7

Conclusions

This thesis work was aimed at generating a new method of analysis of small angle scattering measures of Intrinsically Disordered Proteins, with particular reference to α - synuclein and its four mutants (A53T, A30P, E46K, G51D) in the monomeric phase, whose aggregation's products, called amyloids fiber are responsible for Parkinson's disease. The new method (VBWSAS) developed in this work proved to be able to recognize, with an important sensitivity, the conformational propensities of each amino acid of the protein by associating, with the small angle scattering experimental curve, a weight factor w_i, to possible molecular models. The method was validated through the use of SAS simulated curves, and was then applied to SAXS curves of WT α - syn and of α - syn's mutants recorded at the European Synchrotron of Grenoble. The results obtained open new scenarios on the methods of investigation of IDPs and their aggregation's processes and confirm the SAS technique as important complementary to those already used in the study of IDPs. The results obtained are confirmed by the literature and this puts

the VBWSAS method as a method of data analysis SAS that can be very important in the study of the conformational behaviors of IDPs as it extracts information directly related to the amino acid sequence from the curve. The discovery of this information regarding the propensity as a function of the amino acid sequence could orient the design of drugs with which to attempt to inhibit or reverse the development of the process of neurodegenerative diseases. In the future, using this method, numerous experiments can be performed to clarify the conformational behavior of α - syn and other proteins such as β - amyloid involved in Alzheimer's Disease may be studied.

Chapter 8

Nanoparticles

In the course of my PhD, in addition to the work shown on the Intrinsically Disordered Proteins, I have been involved in some projects concerning Nanoparticles (NPs) and their interaction with proteins or other biological molecules in order to characterize drug-delivery systems.

8.1 Structural and Thermodynamic Properties of Nanoparticle-Protein Complexes: A Combined SAXS and SANS Study

This work [41] is the development of a novel method for determining the structural and thermodynamic properties of NPs - proteins complexes under physiological conditions.

The method consists of collecting a full set of SAXS and SANS measurements in solutions with different concentrations of NPs and proteins. The NPs - proteins dissociation process is described in the framework of the Hill cooperative model, based on which the whole set of X-ray and neutron-scattering data is fitted simultaneously. This method is applied to water solutions of gold - NPs in the presence of human serum albumin without any previous manipulation and can be, in principle, extended to all systems. The work demonstrates that the protein dissociation constant, the Hill coefficient, and the stoichiometry of the NP - protein complex are obtained with a high degree of confidence.

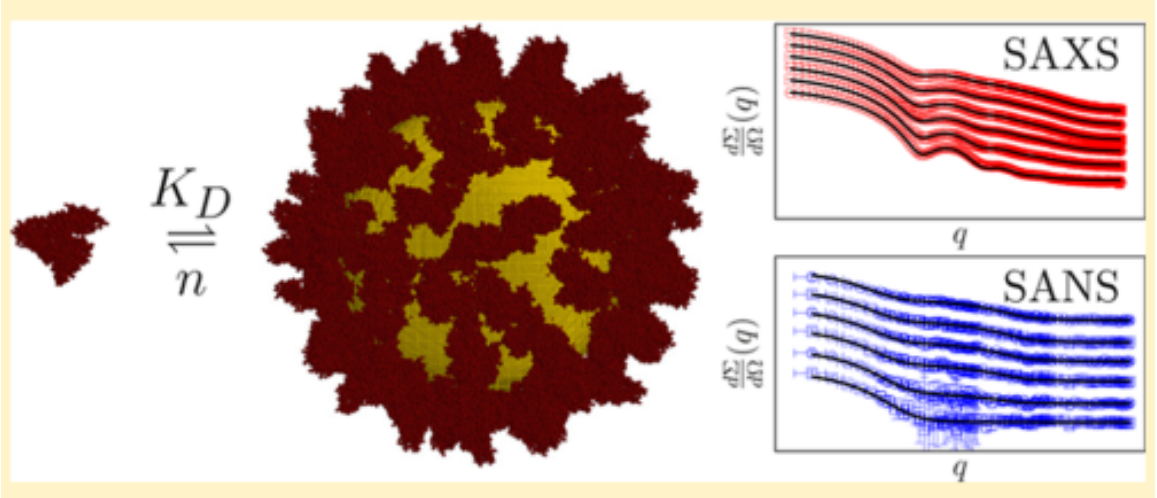


FIGURE 8.1: SAXS-SANS curve of gold nanoparticles in interaction with Human Serum Albumin

8.1.1 Introduction about the system

Engineered nanomaterials (ENMs) are almost routinely used in many consumer products, ranging from health and fitness to energy and food [42]. One of the most promising application areas of ENMs, particularly functionalized NPs is medicine, including both medical treatment and diagnostics[43]. However, although many of the peculiar properties of nanomaterials are deemed to be beneficial, there is still a wealth of concern about the potential risks that ENMs may cause when in contact with humans and more in general with the environment[44]. A great deal of efforts are currently being made to understand the mechanisms governing the interactions between ENMs and cells, with a special focus on engineered NPs. It is generally accepted that when NPs placed in contact with a biological medium (e.g. blood, cell culture, etc.) they are immediately covered by a layer of proteins, known as protein corona (in Fig. 8.2), but there are still several open questions concerning the interaction mechanisms and the influence of different thermodynamic parameters [45–47].

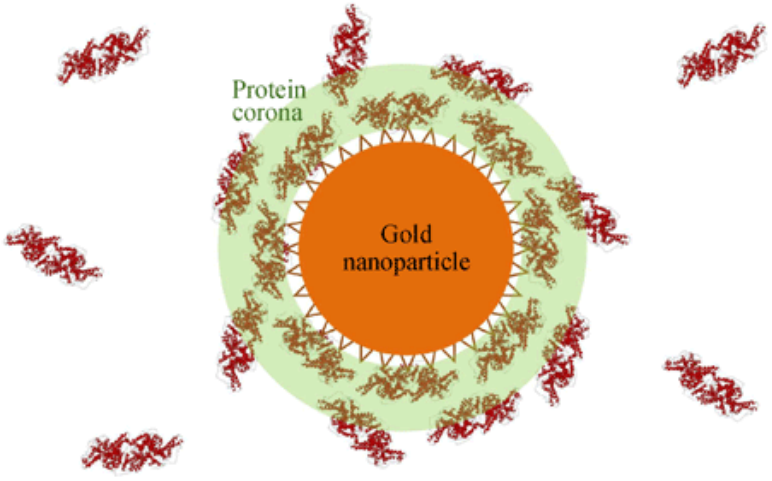


FIGURE 8.2: Gold Nanoparticle structure and protein corona. [18]

In this work, the intriguing mechanism of NP - protein interactions is investigated by concurrently using SAXS and SANS to directly inspect the samples containing both proteins and NPs without any earlier manipulation: this allows for physiological conditions while inspecting the structure and thermodynamics of the NP - protein dissociation. The SAS technique, especially SAXS, has been indeed recently reviewed and recognized to play a pivotal role in the investigation of in situ and in operando NP materials [48]. The approach was as follows:

- We computationally simulated the sensitivity of SAS to detect NP - protein dissociation for gold NPs interacting with human serum albumin (HSA). This allowed the determination of the "optimal" concentration ranges of NPs and protein to perform actual SAXS and SANS experiments.

- Hill thermodynamic model was exploited for simultaneously analyzing several sets of SAS data collected for different concentrations of both proteins and NPs using the GENFIT software[49].

The final results of the best fits are the ΔG° free energy of the NP - protein dissociation process and the mean number of proteins associated with an NP when a protein monolayer is formed, that is, the -called stoichiometry of the NP-protein interaction. The outcome of the whole set of experiments and data elaboration described here is, on the one hand, to obtain important information on the stoichiometry of the NP-protein interaction, and on the other hand, to establish an experimental protocol that can easily be transferred to other systems.

8.1.2 Materials and Methods

The goal of this work was to study form a thermodynamic and structural point of view the interactions of proteins with spherical NPs, constituted by a gold core and coated with a shell of stabilizing materials, such as citrate anions. A new method has been developed in order to understand the main aspects of the dissociation of proteins from stabilized NPs and describe their structure at low resolution, according to the information that can be obtained from a batch of SAXS and SANS experiments conducted on these complexes.

8.1.2.1 Thermodynamic Model

The thermodynamic model used in this method is a dissociation model called Hill model [50]. This model works by assuming that:

- Protein molecules could form a monolayer on an NP;
- A dissociation equilibrium is established between the proteins bound to an NP and free protein molecules in solution;
- The dissociation involves n protein molecules at a time and takes place at identical association sites, the number of which is constant

After the association with protein molecules, the NPs can be seen as polydisperse spheres with an average core radius $< R >$, coated with two layers:

- the first layer, of thickness t_1, is compact and built by the stabilizing molecules;

- the second layer, t_2, is composed of a mixture of water and protein molecules.

The volume fraction filled with proteins in the second layer is variable and depends on the thermodynamic equilibrium of the dissociation of n, the number of protein molecules (L) bound to n association sites (S) above the surface of the first layer, according to the following elementary Hill process:

$$\frac{1}{n}L_nS_n \rightleftharpoons L + \frac{1}{n}S_n \tag{8.1}$$

The Hill coefficient n describes the cooperativity of protein binding:

- if $n > 1$, once a protein is bound to the NP, its affinity for binding other proteins increases;
- for $0 < n < 1$, the affinity to other proteins decreases after binding with the first protein.
- for $n = 1$, there is no cooperativity effect and the sites behave independently, according to the classical Langmuir model [51].

The Hill process (eq. 8.1) is described by the dissociation equilibrium constant:

$$K_D = e^{-\Delta G^\circ / K_B T} = C_L C_{S_n}^{1/n} / C_{L_nS_n}^{1/n} \tag{8.2}$$

which depends on the molar concentrations of n protein-occupied sites ($C_{L_nS_n}$) on the free proteins in solution (C_L) and free n sites (C_{S_n}). The corresponding reference dissociation Gibbs free energy difference is ΔG°, and K_B and T are the Boltzmann constant and the absolute temperature, respectively.

The structural properties of proteins and NPs are included in the thermodynamic model as follows. We consider each protein molecule attached to the NP as forming a "protein box" with an average basis surface a_L and height t_2 . The number of protein association sites ($\$N_{max}$) over the NP is therefore determined by dividing the surface a_{NP} of the polydisperse stabilized NP by a_L . By assuming a lognormal distribution of the NP core radius, the average surface of the stabilized NP is given by:

$$a_{NP} = 4\pi \left[\langle R \rangle^2 (1 + \xi^2) + 2 \langle R \rangle t_1 + t_1^2 \right] \tag{8.3}$$

where ξ is the core radius dispersion, defined as $\xi^2 = \langle R \rangle^{-2} (\langle R^2 \rangle - \langle R \rangle^2)$

we then consider the mass balance equations. For the NPs, the concentration of the total number of sites of association, expressed by $N_{max}C_{NP0}$ is the total concentration of the NPs, should be equal to the sum of the concentrations of free and occupied sites, according to $N_{max}C_{NP0} = n(C_{S_n} + C_{L_nS_n})$.

We notice that C_{NP0} is obtained by dividing the nominal concentration of the metal or oxide forming the core of NPs (indicated by the symbol M) by the number X_M of the M atoms (or molecules) that, on average, are contained in one NP, $C_{NP0} = C_{M0}/X_M$.

The value of X_M is given by the ratio between the average volume of the log-normal polydisperse bare NP, $\nu_{NP} = (4\pi/3)\langle R\rangle^3 (1+\xi^2)^3$, and the molecular volume of an M unit, $\nu_M = M_{wM}/(d_M N_a)$, where M_{wM} and d_M are the molecular weight and the mass density of the M unit, respectively, and N_a is the Avogadro constant. The mass balance for proteins is $C_L + nC_{L_n}C_{S_n} = C_{L0}$, where C_{L0} is the total nominal protein concentration in the sample. Combining the two mass balances with the definition of association constant, the fraction of association sites occupied by protein molecules, defined as $\bar{Y} = C_{L_nS_n} + C_{S_n}$, and the concentration of free protein molecules in solution should satisfy to the following equations:

$$K_D^n = \frac{1-\bar{Y}}{\bar{Y}} \left(C_{L0} - N_{max}\bar{Y}C_{NP0} \right)^n \tag{8.4}$$

$$C_L = C_{L0} - N_{max}\bar{Y}C_{NP0} \tag{8.5}$$

In general eq. 8.4 can be solved only by numerical methods. According to the low-resolution structural model adopted here, $\bar{Y}$ can also be interpreted as the volume fraction of the second layer of the NP occupied by the protein boxes mentioned above. The dry-protein volume fraction in each protein box is $\psi = \nu_L/(a_L t_2)$, where the dry volume ν_L of one protein molecule can be easily calculated by summing up the van der Waals volumes

of all of its atoms. The overall dry protein volume fraction of the second shell is hence the product $\bar{Y}\psi$.

8.1.2.2 Scattering Cross Section

SAS experiments provide the macroscopic differential scattering cross section $\frac{d\Sigma}{d\Omega}(q)$ as a function of the scattering vector **q** as shown in chapter 2. In the case of NPs, in presence of protein molecules, the scattering intensity is a combination of both form factors and partial structure factors. The form factor include on of th NPs, $P_{NP}(q)$, and that of the protein, $P_L(q)$, whose relative weighting factors are dictated by the respective concentrations of NPs (C_{NP0}) and free protein molecules (C_L). The partial structure factor $S_{\alpha\beta}(q)$ are the Fourier transforms of the long-range correlation functions $g_{\alpha\beta}(r)$ between particles of kinds α and β where $\alpha\beta$ stands for NP-NP, NP-L and L-L interactions, and are weighted by $(C_\alpha C_\beta)^{\frac{1}{2}}$ [52]. If the molar concentration of NPs is many orders of magnitude lower than that of proteins, the only partial structure factor that may be significantly different from 1 is that describing the L-L correlation. In these conditions, the scattering intensity can be written in the following form:

$$\frac{d\Sigma}{d\Omega}(q) = N_A \left[C_{NP0} P_{NP}(q) + C_L P_L(q) S(q) \right] + B \tag{8.6}$$

- $P_{NP}(q)$ is obtained by integrating the form factor $P_{3S}(q, R)$ of a three-density-level sphere with a core radius R over a log-normal distribution $f(R)$:

$$P_{NP}(q) = \int dR f(R) P_{3S}(q, R) \tag{8.7}$$

- The form factor $P_L(q)$ of a globular protein, the atomic structure of which is known can be easily calculated from a coordinate file deposited in the PDB - Protein Data Bank [53] using SASMOL method [15] included in GENFIT: the positions of the hydration water molecules are found by embedding the protein molecule in a tetrahedral close - packed lattice that mimics the structure of liquid water. Form factors are calculated as function of the kind of radiation exploited (X-ray or Neutrons), the exchangeable deuteration grade X_D (in the case of SANS), and the relative mass density d_w of the hydration water molecules surrounding the free protein molecules in solution;

- The term $S(q)$ is the effective structure factor defined as:

$$S(q) = 1 + \frac{[\ F_L(q)]^{\ 2}}{P_L(q)} [\ S_{LL}(q) - 1] \tag{8.8}$$

where $F_L(q)$ is the orientational average of the scattering amplitude of a protein molecule (determine by the SASMOL method) and $S_{LL}(q)$ is the protein-protein structure factor. The function $S_{LL}(q)$ implemented in GENFIT is described in terms of the mean spherical approximation by taking into account a screening Coulumbian

repulsive potential and a Yukawian attractive potential via random phase approximation.

The parameters of this model are the average protein diameter D, the number z of protein net elementary charges (e), the added ionic strength I_s, the attractive energy J of two proteins in contact, and the range d of the mutually attractive potential. The additional term B in eq.8.8 is a flat background, which takes into account incoherent scattering effects. Equation 8.8 deserve an important comment. When SAXS or SANS curves are recorded by varying only the nominal concentration of NPs or proteins (C_{NP0} and C_{L0}, respectively, eq.8.5), the whole set of data can be fitted by 8.8 using a unique analysis, the main fitting parameters of which are the reference dissociation Gibbs free energy difference ΔG° and the cooperative index n, which, together with the values of nominal concentrations, determines the value of $\bar{Y}$, according to eq.8.4 and 8.5. On the other hand, $\bar{Y}$ is the main variable that establishes the scattering length density of the second shell. To make the global fit analysis of data more robust and to validate the suitability of the used PDB file, it is convenient to perform SAS measurement on proteins without NPs at different concentrations and to include the resultant data in the global fit analysis. The best-fit parameters of the protein form factor $P_L(q)$ are shared by all SAS curve, with and without NPs.

8.1.2.3 Result and Discussion

The best-fit results of all SAS data are shown in Fig. 8.3.

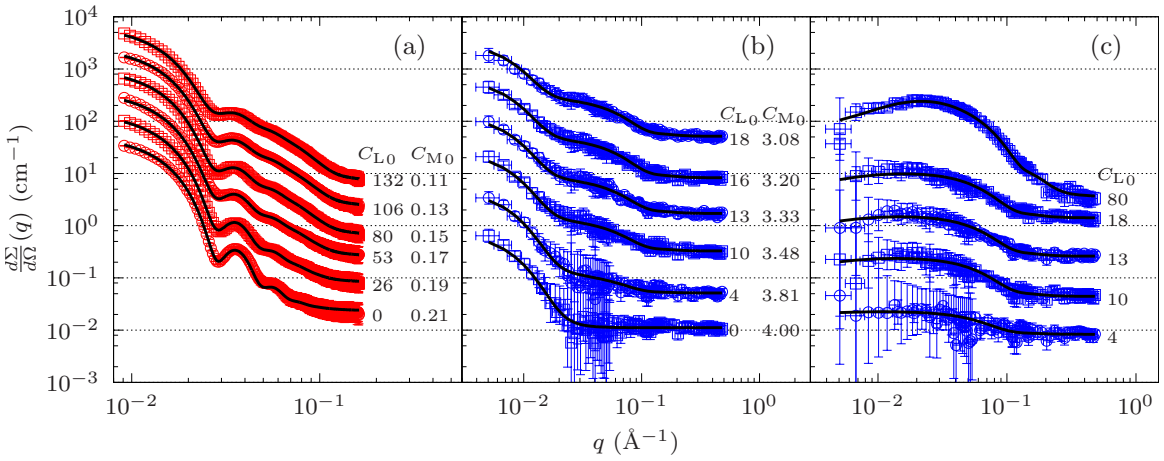


FIGURE 8.3: Experimental and fitted SAXS and SANS data of gold NPs in the presence of HSA. Total molar concentrations of HSA, C_{L0} , and gold, C_{M0} ,are reported next to each curve in micromolar and millimolar units, respectively. The solid black lines are the best-fit curves obtained by GENFIT.

(a) SAXS data of NPs in the presence of HSA recorded at the European Synchrotron Radiation Facility. The curves are stacked by a factor of 3 for clarity.

(b) SANS data of D_2O solutions of gold NPs in the presence of HSA. The curves are stacked by a factor of 5, for clarity.

(c) SANS data of HSA in D_2O solution recorded at the Laboratoire Leon-Brillouin. The curves are stacked by a factor of 5, for clarity. The length of the horizontal bar shown for each q of all SANS data represents twice the experimental standard deviation σ_q .

All fitting parameter, apart from the background B, are reported in Table 8.1 with their standard deviations.

Common Fitting Parameters Obtained for the SAXS and SANS Curves Reported in Fig.8.3

χ^2	a		1.04
ΔG°		(K_BT)	11.0 ± 0.5
K_D	a	(μM)	17 ± 8
n			0.8 ± 0.1
$\langle R \rangle$	b	$\mathring{A}$	154.6 ± 0.2
ξ	b		0.101 ± 0.002
$\langle R \rangle$	c	$(\mathring{A})$	145 ± 9
ξ	c		0.35 ± 0.04
a_L		$(\mathring{A}^2)$	3200 ± 400
N_{max}	a, b		100 ± 20
N_{max}	a, c		100 ± 30
t_2		(Å)	30 ± 3
X_{DM}			0.981 ± 0.006
d_w			1.081 ± 0.0007
D		(Å)	63 ± 4
z		(e)	-4.2 ± 0.2
I_s		(μM)	50 ± 20
J		(K_BT)	30 ± 1
d		(Å)	3.1 ± 0.4

TABLE 8.1: (a) Derived parameters. (b) Parameters of SAXS curve. (c) Parameters of SANS curve.

The quality of data fitting appears to be quite high, with an average χ^2 as large as 1.04. Clearly, the most relevant parameters are the unique values of the free energy difference ΔG° that assists the dissociation process of one HSA molecule from the NP surface and the Hill cooperative index n.

The small value $\Delta G^\circ/(K_B T) = 11.0 \pm 0.5$ means that the corona formed by HSA on gold NPs is a *soft corona*, according to Lundqvist Et Al.[54]. The dissociation constant related to ΔG° is found to be $K_D = 17 \pm 8 \mu M$. The cooperative index obtained is significantly lower than 1 ($n = 0.8 \pm 0.1$),indicating, in agreement with literature [55][56] a certain degree of steric hindrance to the binding of other protein molecules to the NP surface after the binding of the first ones. The area occupied by a single HSA molecule when associated with the gold NP, $a_L = 3200 \pm 400$ Å^2, and the protein layer thickness, $t_2 = 30 \pm 3$ Å, show that the HSA-NP interface mainly involves the basis of the triangular prism approximating the protein shape. For both SAXS and SANS series, the average NP radii are very close to the nominal value of 150 Å; however, a quite high difference is stated in their log-normal dispersion ξ, that is, 0.101 ± 0.0002 for SAXS and 0.35 ± 0.04 for SANS. We can attribute this difference, which is, however, in agreement with the size distribution obtained using transmission electron microscopy, dynamic light scattering and collective light scattering to possible batch-to-batch variations. Among the derived parameters, focus should be brought on the maximum number of protein site N_{max}, which represents a higher limit for the stoichiometry of the NP-protein association: similar results for all SAXS and SANS patterns were obtained, that is, 100 ± 20 and 100 ± 30, respectively. The trend of the fraction of association sites occupied by protein molecules,$\bar{Y}$, is plotted in Fig. 8.4 versus the free protein in solution,C_L, after model

fitting in the standard form (left frame) and as a Hill-plot form (right frame).

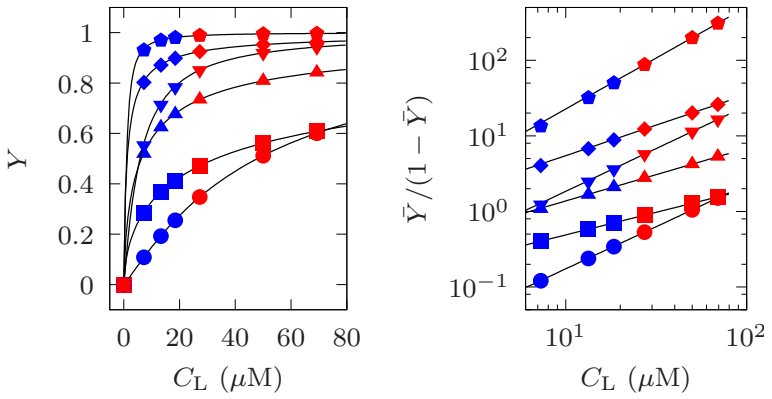


FIGURE 8.4: Left frame: Fraction of the association sites occupied by protein molecules obtained by the global fit analysis of experimental SAXS and SANS curves for gold NPs in the presence of HSA plotted as a function of the in-solution free protein concentration. Red and blue circles refer to SAXS and SANS data, respectively. The solid black lines feature the trend calculated via eq. 8.4 and 8.5 for values of ΔG° and n obtained by data fitting. The dashed black lines represent the trend for the same value of ΔG° and for $n = 1$. Right frame: the same data as a Hill plot.

The data points, with the derived error bars, refer to different ranges of protein concentration probed via SAXS (red points) and SANS (blue points). The solid black lines are the theoretical solutions of eqs 8.4 and 8.5 with the numerically optimized parameters. The dashed line describes the trend of $\bar{Y}$ with the same values of ΔG° obtained through the fit but with $n = 1$. Within the derived error bars, it is clear that from the experimental framework of the presented SAS data, estimating a cooperative index lower than 1 can be seen as a robust achievement. The best - fit parameters relevant to the protein - protein interaction in solution (effective protein diameter D, protein net charge z, added ionic strength I_s , and depth and range of the attractive potential, J and d, respectively) are in good agreement with the literature results of HSA in neutral pH [57]. As final remarks, we wish to underline that the stability of the fitted parameters has required SAS measurements in absolute units and that the nominal concentrations of gold and

HSA to be kept fixed to the values originating from the sample preparations.

In conclusion, in this work it has been shown that whole sets of SAXS and SANS data can be simultaneously processed for studying the structure and thermodynamics of proteins associated with NP surfaces. The thermodynamics of the process is characterized by the ΔG° free energy and the data obtained from the data best fits suggests the formation of a **soft corona** and the n parameter of the Hill model, a figure of 0.8 ± 0.1 was obtained, clearly indicating that the adsorption of one protein molecule on the NP surface moderately depresses the affinity for the associations of other protein molecules. The structure of the NP interacting with HSA is well modelled by a three - density - level sphere. The inner core is the Au - NP, being covered by citrate anions while the outer layer is a monolayer of HSA: we found that for 300 - Å - Au NPs, the formation of the HSA monolayer is consistent with the association of ≈ 100 molecules. SANS and SAXS data in this work were collected under physiological conditions and this paves the way for in situ and in operando studies of the complex problem of NP-protein association.

8.2 A Polaxamer-407 modified Liposome encapsulating epigallocatechin-3-gallate in presence of magnesium:Characterization and protective effect against oxidative damage

This work [58] concern the characterization of a drug delivery system in which the epigallocatechin-3-gallate (Fig 8.5) has been encapsulated.

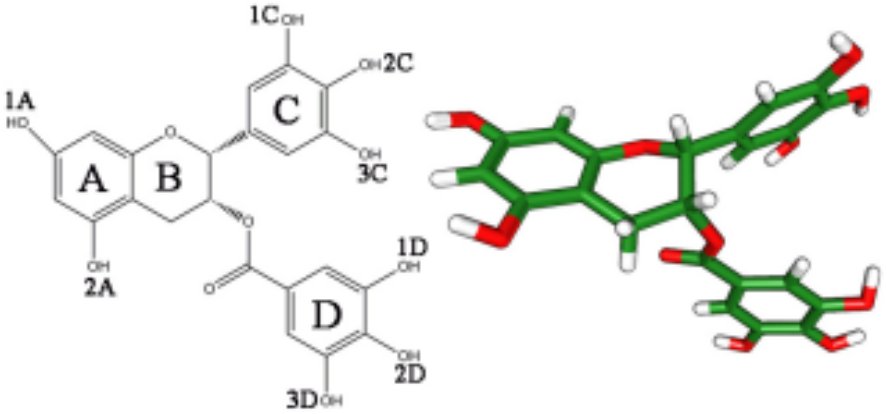


FIGURE 8.5: Structure of epigallocatechin-3-gallate (ECG)

EGCG has been encapsulated inside anionic liposomes made of 1-palmitoyl-2-oley-sn-glycero-3-phosphocholine, and cholesteryl hemisuccinate. To maximize EGCG internalization, magnesium salt was added in the preparation. However stable nanodispersions suitable for drug delivery were obtained only after treatment with Poloxamer-407, a polyethylene-propylene glycol copolymer. The structural and morphological properties of the produced dispersion were studied by X-ray diffraction, which showed a multilamellar structure even after EGCG addition and an ordering effect of Poloxamer-407; Dynamic

Light Scattering demonstrated serum stability of the liposomes. The characterization was completed by evaluating both encapsulation efficiency (100% in the final formulation) and in vitro EGCG release. Since oxidative stress is involved in numerous retinal degenerative diseases, such as age-related macular degeneration, the ability of these liposomes to contrast H_2O_2 - induced cell death was assessed in human retinal cells. Morphological changes at the subcellular level were analyzed by Transmission Electron Microscopy, which showed that mitochondria were better preserved in cells treated with liposomes then those treated with free EGCG.

In conclusion, the results demonstrated that the produced formulation enhances the efficacy of EGCG under stress conditions, thus representing a potential formulation for the intracellular delivery of EGCG in diseases caused by oxidative damage. In this thesis I present a brief description of the drug delivery system, and the part that involved me, the characterization through X-ray Diffraction technique of the system.

8.2.1 Introduction about the drug delivery system

Oxidative stress can be linked to several pathophysiological processes including retinal degenerative diseases, such as age-related macular degeneration (AMD) where an increase in the steady-state concentration of reactive oxygen species (ROS) is present [59].A possible strategy to contrast oxidative stress is to reduce the concentration of ROS, like superoxide anion, hydrogen peroxide,and hydroxyl radical, by enhancing the level of antioxidant molecules in the tissues. Amongst the water-soluble antioxidants, catechins, the main polyphenolic compounds present in green tea, show considerable bioactivity

in degenerative diseases associated with oxidative stress [60]. Epigallocatechin-3-gallate (EGCG), the major constituent in green tea, is mainly responsible for the remarkable antioxidant activity due to the presence of the D ring in the galloyl group in addition to the other three rings present in its structure (A, B and C) (Fig. 8.5) which are sensitive to oxidation [61]. These structural characteristics account for EGCG's potent radical scavenging activity [62] towards both superoxide and hydroxyl radicals as well as peroxyl radicals, nitric oxide, carbon-centered free radicals and lipid free radicals.

Among all drug delivery systems, liposomes offer several advantages because of their biocompatibility, their low toxicity and non-immunogenicity [63]. These supramolecular aggregates can encapsulate enzymatic antioxidants but also hydrophilic and lipophilic chemical antioxidants, shielding and protecting them from inactivation or rapid clearance from cells. Moreover, the ability of polyphenols to interact with the lipid bilayer [64] promotes the encapsulation of these compounds inside lipidic nanoparticles and makes liposomes potential delivery systems for EGCG [65]. In particular, by using anionic multilamellar liposomes prepared from a ternary, lipidic system (POPC/ DOPE /CHEMS 1:1:1)(Fig. 8.6) in the presence of Mg^{2+} ions, we obtained complete EGCG encapsulation. However, the liposomes resulted very unstable, as the simultaneous presence of EGCG and magnesium salts was observed to induce the formation of large cluster aggregates. In order to obtain nanoparticles useful for EGCG delivery applications in this study we describe the preparation and characterization of liposomes made from the same ternary lipidic system, containing EGCG and magnesium salts in addition to a stabilizing agent (the Poloxamer-407) which is able to prevent aggregation and to promote the formation of stable nanodispersions.

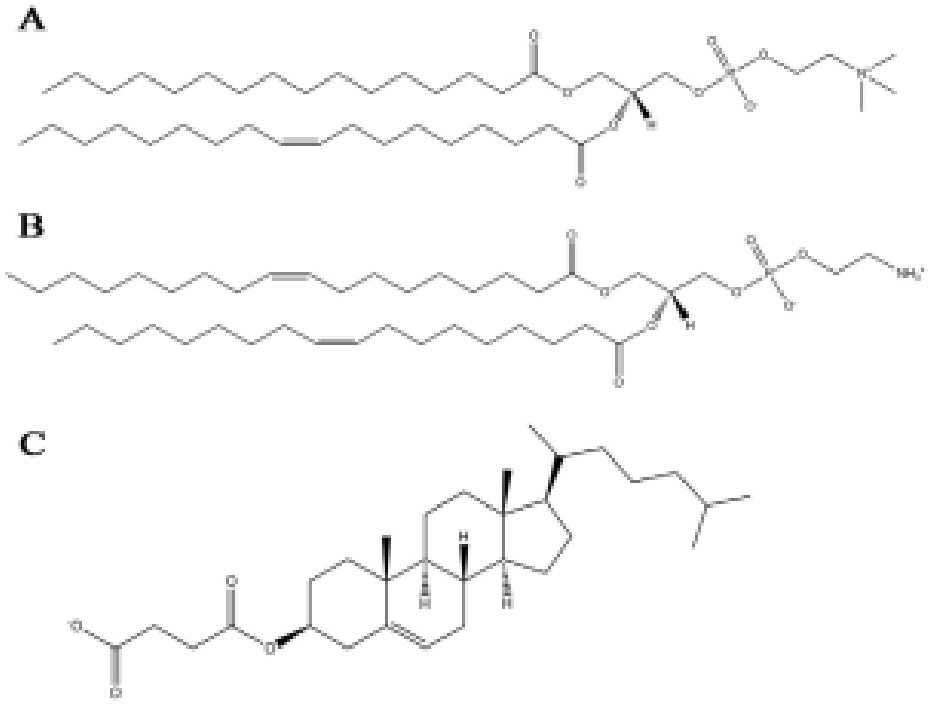


FIGURE 8.6: 2D structure of POPC (A), DOPE(B) and CHEMS(C).

It is noteworthy that Poloxamer-407 is a non-ionic triblock copolymer composed by a central hydrophobic poly(propylene oxide) chain (PPO) capped by two hydrophilic chains of poly(ethylene oxide) (PEO), which functions as emulsifier and stabilizer [66]. Poloxamer-407 is approved as an inactive ingredient by the FDA for various types of pharmaceutical formulations, and is accepted as GRAS (Generally Recognized as Safe) excipient [67], making it an excellent candidate for the preparation of drug delivery systems.

8.2.2 Material and Methods

8.2.2.1 Preparation of bulk and nanodispersed liposomal phases

The preparation of bulk and nanodispersed liposomal phases was done by the chemistry group of the Department of Life Sciences at the Università Politecnica delle Marche. The main characteristics of the protocol are shown below.

Bulk liposomal phases were obtained by Reverse Phase Evaporation (REV)[68]. Appropriate amounts of chloroform solutions of DOPE, POPC, CHEMS and methanol solution of EGCG, when present, were mixed to obtain $1:1:1:0.6$ mol/mol ratio and a final concentration of 3 mg/mL^{-1} of lipids and 1 mM of EGCG. The solvent was removed under reduced pressure at room temperature to preserve the EGCG molecular structure [69]. After removal of residual solvent under nitrogen flow, lipids were redissolved in 3 mL of an ether/methanol mixture $(2:1, v/v)$ and $1mL$ of phosphate buffered saline (PBS, pH 7.4) was added with or without $MgCl_2$ salts ($MgCl_2/EGCG, 5:1$ mol/mol). With the aim to obtain an initial water-in- oil emulsion (W/O), the resulting two-phase system was briefly sonicated (2 min) with a vibra cell sonicator (Sonics Vibra Cell Mod.VCx130) equipped with a tapered micro tip. The organic solvent was removed under vacuum (Rotavapor, Büchi) to cause a phase inversion that gave an O/W emulsion. The obtained liposomes L, ML (magnesium containing liposomes), L-EGCG (EGCG loaded liposomes) and ML-EGCG (magnesium containing liposomes loaded with EGCG) were characterized fresh and/or after equilibration for 24 h. The liposomal suspensions containing Poloxamer-407, PxL (poloxamer liposomes), MPxL (magnesium-containing poloxamer liposomes), PxL-EGCG (poloxamer liposomes loaded with EGCG) and MPxL-EGCG

(magnesium-containing poloxamer liposomes loaded with EGCG) were prepared in the same way, but Poloxamer-407 was added in PBS to obtain a polymer final concentration of 0.8 $mg\ mL^{-1}$. Note that the MLV suspensions were directly used for X-ray diffraction experiments, while samples for DLS characterization, turbidimetric analysis, encapsulation efficiency determination, in vitro release and cellular assays, were sonicated (sonic Vibracell) before being used for 30 min in pulse mode (30 s on; 2 s off,50%) at 0° C, until the liposome dispersion was completely clear.

8.2.2.2 X-ray diffraction

X-ray diffraction experiments were performed using a 3.5 kW Philips PW 1830 X-ray generator (Amsterdam, Netherlands) provided with a bent quartz crystal monochromator ($\lambda = 1.54\,\text{Å}$) and a Guinier-type focusing camera (homemade design and construction, Ancona,Italy). Diffraction patterns were recorded on GNR Analytical Instruments Imaging Plate system (Novara, Italy). MLV suspensions were measured in a tight vacuum cylindrical cell equipped with thin mylar windows. Experiments were performed as a function of temperature, at 25, 36, 40 and 45°C. In each experiment a few Bragg peaks were detected. Peak indexing was performed considering the usually observed lipidic phases [70] and the unit cell dimension of the phases, d, calculated from the averaged spacing of the observed peaks.According to the decomposition of the sample in the hydrophobic and hydrophilic regions [71], a simple equation relates the unit cell dimension to the lipid head-to-head distance, d_{HH} , that can be measured from electron density maps [72], and the thickness of the water layer d_w :

$$d_w = d - d_{HH} \tag{8.9}$$

Therefore, the area-per-lipid at the water/lipid interface S_{lip} and the averaged number of water molecules associated with one lipid molecule n_w/lip can be determined if v_{lip} and v_{wat}, the averaged lipid molecular volume (in this case estimated by molecular dynamics simulations to be around 960 Å^3) and the water molecular volume (30 Å^3) are known:

$$S_{lip} = \frac{2v_{lip}}{d_{HH}} \tag{8.10}$$

and

$$n_{w/lip} = \frac{V_{w/lip}}{v_{wat}} \tag{8.11}$$

with $V_{w/lip} = d_w S_{lip}/2$.

8.2.3 Results and Discussion.

To analyze the structure of the MLVs and the influence of EGCG and magnesium cations on the supramolecular properties of the self-assembled nanoparticles, X-ray diffraction experiments were carried out on L, ML, L-EGCG and ML-EGCG systems, prepared

both in the absence and in the presence of Poloxamer-407. The X-ray diffraction profiles obtained in the absence of Poloxamer-407 are reported in Fig. 8.7.

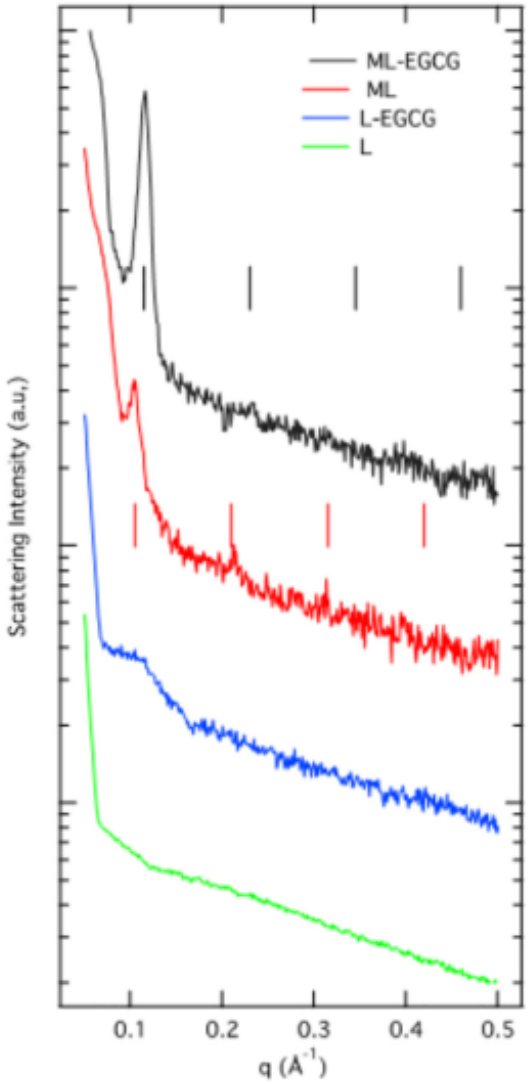


FIGURE 8.7: X-ray diffraction profiles from L,ML, ML-EGCG, L-EGCG systems at 25°C

Depending on composition, different characteristics are observed: for L and L-EGCG samples, a very large band (centered at about 57 Å) occurs in the low-Q region, suggesting the formation of a disordered system, probably unordered MLVs or LUVs; on

the other side, 3 or 4 Bragg peaks characterize the diffraction profiles of ML and ML-EGCG samples, indicating that the presence of Mg^{2+} induces the formation of a rather ordered structure. As the peak spacing ratios scale as 1:2:3..., a 1-D lamellar organization is proposed: the well-known rearrangement of the lipidic carbonyl region [73], which accompanies Mg^{2+} binding and which involves hydration and conformational changes, appears to stabilize the lamellar structure, even in the presence of EGCG. ML and ML-EGCG samples then show the same multilamellar structure, the main difference being the unit cell dimension (e.g., the repeat distance among the lamellar, d), which is larger for ML (59.8 ± 0.5 Å) and smaller for ML-EGCG (54.6 ± 0.5 Å). The 10% reduction observed in the presence of EGCG can be related to a different lipid hydration induced by the presence of the active molecule or to changes in the hydrocarbon chain conformation, which is more disordered when EGCG is present. To disentangle the different structural effects, low-resolution electron density profiles have been calculated, as reported by [72]. Electron density maps, reported in Fig 8.8, show small differences in the position of the maxima corresponding to the head group location and in the form of the electron density in the hydrocarbon region (see the differences between the full and the dashed lines).

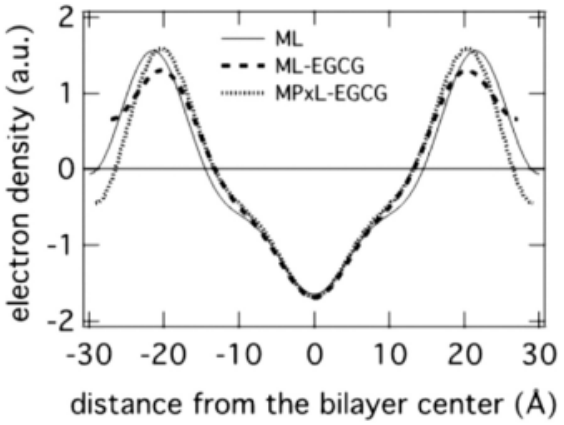


FIGURE 8.8: Reconstructed electron density profiles for the lamellar phase of ML (full line), ML-EGCG (dashed line) and MPxL-EGCG (dotted line) systems at 25°C.

From the head-to-head electron density peak distance, the lipid bilayer thickness was obtained and, according to the decomposition of the sample in the hydrophobic and hydrophilic regions [71] the average lipid cross sectional area, S_{lip} , and the lipid-associated intermembrane water volume, $V_{w,lip}$, were derived using equations 8.10 and 8.11. The results shown in Table 8.2 are compatible with the location of EGCG near the lipid head-group, so as to induce an increase in the area-per-lipid at the polar/apolar interface and an increased disorder of the lipid hydrocarbon chains. Of note is that hydration is not modified, as indicated by the rather constant lipid-associated intermembrane water volume (e.g., the number of water molecule-per-lipid is practically the same); consequently the formation of aggregates has to be ascribed to the strong interaction between EGCG and lipid-head group rather than to dehydration phenomena.

System	**d (Å)**	$\mathbf{d}_{HH}$(Å)	$\mathbf{d}_{wat}$(Å)	$\mathbf{S}_{lip}$(Å^2)	$\mathbf{V}_{w/lip}$(Å^3)	$\mathbf{n}_{w/lip}$
ML	59.8	44	15.8	43.6	345	11
ML-EGCG	54.6	40	14.6	48.0	350	12
MPxL-EGCG	58.1	40	18.2	48.0	436	-

TABLE 8.2: Structural and molecular parameters for ML, ML-EGCG and MPxL-EGCG.

This result can be related to the high encapsulation efficiency observed in our liposomal system but also to the occurrence of precipitates in the liposome formulation containing both Mg^{2+} and EGCG. Indeed, the interaction network is complex. On the one hand, EGCG can behave both as H-donor with lipid oxygen groups and as H-acceptor with ethanolamine groups in DOPE; on the other hand, the presence of magnesium ions enhances the interaction possibilities between EGCG and the lipid bilayer. In fact, at physiological pH, EGCG is partially deprotonated (its pKa is 7.75) so that magnesium divalent ions can serve as cross-bridge to bind phosphate or oxygen lipid groups. Moreover, the presence of CHEMS, which is a negatively charged steroid molecule, increases the electrostatic interactions with Mg^{2+} ions, indirectly promoting the interaction of EGCG with the lipid bilayer [64] X-ray diffraction profiles related to samples prepared in the presence or absence of Poloxamer-407 were very similar. Fig. 8.9 shows the results obtained from ML-EGCG and MPxL-EGCG carriers as a function of temperature: in both cases the multilamellar structure is confirmed by the Bragg peak sequence. Furthermore, the peak reciprocal intensities are maintained even at $45°C$, so guaranteeing the antioxidant liposome stability at physiological temperature.

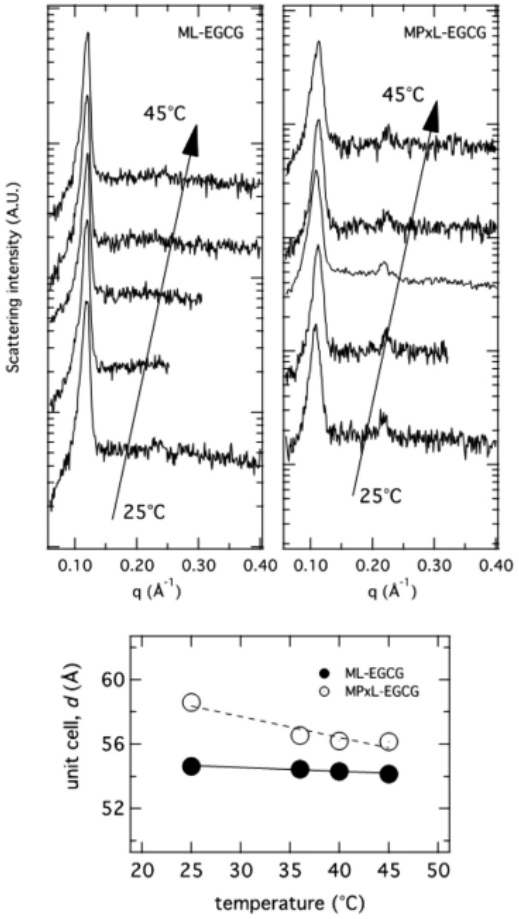


FIGURE 8.9: X-ray diffraction profiles observed for ML-EGCG (left frame) and MPxL-EGCG (right frame) samples as a function of temperature, as indicated. The lower frame shows the temperature-dependence of the lamellar repeat distance, as determined from the analysis of the position of the observed Bragg peaks.

If the two carriers show the same structural stability on heating, the polymer has an ordering effect on the supramolecular organization of liposomes and induces a change in the unit cell, which increases from 54.6 to about 58 Å. According to the unit cell temperature dependence reported in Fig. 8.9, such an increase is probably determined by the presence of polymer molecules in between the lipid layers. In fact, as reported in [74], Poloxamer-407 has two water soluble PEO chains and a more hydrophobic middle block that may be adsorbed at, or incorporated in, the surface of the lipidic structural elements. As a consequence, steric repulsion between bilayers could occur. Electron density

maps confirm this view: the profiles calculated at $25°C$ for ML-EGCG and MPxL-EGCG samples, and reported in Fig. 8.8 (compare the dashed and the dotted lines), show that the head-to-head electron density peak distance is practically unaltered by the presence of Poloxamer-407 (hence, no changes in lipid conformation and in average lipid cross sectional area are expected). Moreover, while the inner lamellar part shows very similar features, differences are observed in correspondence of the polar region. Poloxamer-407 does not affect the lipid packing, but steric effects due to adsorption at the polar-head region enlarge the lamellar repeat distance (note the consequent increases of the lipid-associated intermembrane volume, referred as $V_{w,lip}$, in Table 8.2). In order to verify the time stability, structural studies were also performed on MPxL-EGCG samples after storage at $4°C$ for 1 month: X-ray diffraction profiles confirm that the inner multilamellar morphology is preserved, with a very small reduction of the unit cell dimension to 57.1 Å, well inside the experimental error (estimated around 0.5 Å). In conclusion XRD experiments show the location of EGCG near the lipid head-group and a more ordered lamellar structure obtained by incorporation of Poloxamer-407.

Bibliography

[1] L. Breydo, J.W. Wu, and V.N. Uversky. α-synuclein misfolding and parkinson's disease. *Biochimica et Byophysica Acta 1822*, pages 261–285, 2012.

[2] R. Linding, J. Schymkowitz, F. Rousseau, F. Diella, and L. Serrano. A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.*, 342:345–353, 2004.

[3] V. N. Uversky. Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta*, 1834:932–951, 2013.

[4] M. Yu. Lobanov and O. V. Galzitskaya. How common is disorder? occurence of disordered residues in four domains of life. *Int. J. Mol. Sci.*, 16:19490–19507, 2015.

[5] A.K. Dunker, C.J. Oldfield, J. Meng, P. Romero, J.Y. Yang, J.W. Chen, V. Vacic, Z. Obradovic, and V.N. Uversky. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*, 9:(Suppl2):S1, 2008.

[6] R. Van der Lee, M.Buljan, B.Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Goughand J. Gsponer, D.T. Jones, P.M. Kim, R.W. Kriwacki, C.J. Oldfield, R.V. Pappu, P. Tompa, V.N. Uversky, P.E. Wright, and M.M. Babu.

Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, 114:(6589–6631, 2014.

[7] V.N. Uversky and A.K. Dunker. Understanding protein non - folding. *Biochimica et Biophysica Acta*, 1804:1231–1264, 2010.

[8] R.E Dorsey B.B. Bloem. The parkinson pandemic - a call to action. *JAMA Neurology*, 75:9–10, 2018.

[9] R.E. Dorsey. The parkinson pandemic - a call to action. *Molecules*, 20:2510–2528, 2012.

[10] Gallegos S. et. al. Features of alpha-synuclein that could explain the progression and irreversibility of parkinson's disease. *Cold Spring Harb Perspect Med*, 9:1–10, 2015.

[11] P. Flagmeier, G.Meisl, M. Vendruscolo, T.P. Knowles, C.M.Dobson, A.K. Buell, and C. Galvagnion. Mutations associated with familial parkinson's disease alter the initiation and amplification steps of α-synuclein aggregation. *PNAS*, 113(37): 10328–10333, 2016.

[12] M. Emanuele and E. Chieregatti. Mechanism of α - synuclein action on neurotransmission: Cell-autonomous and non - cell autonomous role. *Biomolecule*, 5(2): 865–892, 2015.

[13] H.D. Mertens and D. Svergun. Structural characterization of proteins and complexes using small angle x-ray solution scattering. *Journal of Structural Biology*, 172:128–141, 2010.

[14] L.R.S. Barbosa, F. Spinozzi, P. Mariani, and R. Itri. Small-angle x-ray scattering applied to proteins in solution. *Proteins in Solution and at Interfaces: Methods and Applications in Bioechnology and Materials Science*, 1:49–72, 2013.

[15] M. G. Ortore, F. Spinozzi, P. Mariani, A. Paciaroni, L. R. S. Barbosa, H. Amenitsch, M. Steinhart, J. Ollivier, and D. Russo. Combining structure and dynamics: non-denaturing high-pressure effect on lysozyme in solution. *J. R. Soc. Interface*, 6: S619–S634, 2009.

[16] T.N. Cordeiro, F.Herranz-Trillo, A. Urbanek, A. Estana, J. Cortes, N. Sibille, and P. Bernardò. Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Current Opinion in Structural Biology*, 42:15–23, 2017.

[17] T. Gurry, O. Ullman, C.K. Fisher, I. Perovic, T. Pochapsky, and C.M. Stultz. The dynamic structure of α-synuclein multimers. *J. Am. Chem. Soc.*, 135:3865–3872, 2013.

[18] M. Cui, R. Liu, Z. Deng, G. Ge, Y. Liu, and L. Xie. Quantitative study of protein coronas on gold nanoparticles with different surface modifications. *Nano Research*, 7:345–352, 2014.

[19] V. N. Uversky. Natively unfolded proteins: A point where biology waits physics. *Protein Science*, 11:739–756, 2002.

[20] V. N. Uversky. Intrinsically disordered proteins from a to z. *The International Journal of Biochemistry and Cell Biology*, 43:1090–1103, 2011.

[21] S. Wallin. Intrinsically disordered proteins: structural and functional dynamics. *Research and Reports in Biology*, 8:7–16, 2017.

[22] P. Tompa. The functional benefits of protein disorder. *J. Mol. Struct. Theochem*, 666-667:361–371, 2003.

[23] P. Tompa. Intrisically unstructured proteins. *Chem. Rev.*, 27:527–533, 2002.

[24] V.N. Uversky, C.J. Oldfield, and A.K. Dunker. Intrinsically disordered proteins in human diseases: Introducing the d^2 concept. *Annu. Rev. Biophys.*, 37:215–246, 2008.

[25] J. Jankovic. Parkinson disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry*, 79:368–376, 2008.

[26] L. Stefanis. α-synuclein in parkinson's disease. *Cold Spring Harb Perspect Med*, 4: a009399, 2012.

[27] J. Li V.N. Uversky and A.L Fin. Effect of familial parkinson's disease point mutation a30p and a53t on the structural properties, aggregation, and fibrillation of human α - synuclein. *Biochemistry*, 40:11604–11613, 2001.

[28] R.A. Frendenburg, C. Rospigliosi, R.K. Meray, J.C. Kessler, H.A. Lashuel, D. Eliezer, and P.T. Lansbury Jr. The impact of the e46k on the properties of α - synuclein in its monomeric and oligomeric states. *Biochemistry*, 46:7107–7118, 2007.

[29] D. Svergun, C. Barberato, and M. H. J. Koch. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.*, 28:768–773, 1995.

[30] F. Merzel and J.C Smith. Sasim: A method for calculating small angle scattering from explicit-atom models. *Acta Cryst. D*, 58:242–249, 2002.

[31] H.J.Dyson and P.E. Wright. Unfolded proteins and protein folding studied by nmr. *Chem. Rev.*, 104:3607–3622, 2004.

[32] C.D. Putnam, M. Hammel, G.L.Hura, and J.A. Tainer. X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quart Rev Biophys*, 40:191–285, 2007.

[33] V. Ozenne, F. Bauer, L. Salmon, J.R. Huang, M.R. Jensen, S. Segard, P. Bernardò, C. Charavay, and M. Blackledge. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, 28:1463–1470, 2012.

[34] R.B. Best, W. Zheng, and J. Mittal. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J Chem Theory Comput*, 10:5113–5124, 2014.

[35] J.Henriques, C. Cragnell, and M. Skepo. Molecular dynamic simulations of intrinsically disordered proteins: Force field evaluation and comparison with experiment. *J Chem Theory Comput*, 11:3420–3431, 2015.

[36] G.N.Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J.Mol.Biol.*, 7:95–9, 1963.

[37] S.C. Lovell, I.W. Davis, W.B. Arendall, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, and D.C. Richardson. Structural validation by $c\alpha$ geometry: ϕ, ψ *and* $c\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50:437–450, 2003.

[38] Charles K. Fisher, Orly Ullman, and Collin M. Stultz. Comparative studies of disordered proteins with similar sequences: Application to aβ40 and aβ42. *Biophys. J.*, 104(7):1546–1555, 2013.

[39] C. L. P. Oliveira, M. A. Behrens, J. S. Pedersen, K. Erlacher, D. Otzen, and J. S. Pedersen. A SAXS Study of Glucagon Fibrillation. *J. Mol. Biol.*, 387:147–161, 2009.

[40] F. Spinozzi P. Mariani and M. G. Ortore. Proteins in binary solvents. *Biophys. Rev.*, 8:87–106, 2016.

[41] F. Spinozzi, G. Ceccone, P. Moretti, G. Campanella, C. ferrero, S. Combet, I. Ojea-Jimenez, and P. Ghigna. Structural and thermodynamic properties of nanoparticle-protein complexes: A combine saxs and sans study. *Langmiur*, 33:2248–2256, 2017.

[42] M.E. Vance, T. Kuiken, E.P. Vejerano, S.P.McGinnis, M.F Jr Hochella MF, D. Rejeski, and M.S. Hull. Nanotechnology in the real world: Redeveloping the nanomaterial consumer products inventory. *Beilstein J. Nanotechnol.*, 6:1769–1780, 2015.

[43] R. Liu, W. Jiang, C. D. Walkey, W. C. W. Chancde, and Y. Cohen. Prediction ofnanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale*, 7:9664–9675, 2015.

[44] S. Sharifi, S. Behzadi, S. Laurent, M.L. Forrest, P. Stroeve, and M. Mahmoudi. Toxicity of nanomaterials. *Chem. Soc. Rev.*, 41:2323–2343, 2012.

[45] I. Lynch, A. Salvati, and K.A. Dawson. Protein-nanoparticle interactions: What does the cell see? *Nat. Nanotechnol.*, 4:546–547, 2009.

[46] F. Turci, E. Ghibaudi, M. Colonna, B. Boscolo, I. Fenoglio, and B. Fubini. An integrated approach to the study of the interaction between proteins and nanoparticles. *Langmiur*, 26:8336–8346, 2010.

[47] E. Casals, T. Pfaller, A. Duschi, G.J. Oostingh, and V. Puntes. Time evolution of the nanoparticle protein corona. *ACS Nano*, 4:2992–3001, 2010.

[48] T. Li, A.J. Senesi, and B. Lee. Small angle x-ray scattering for nanoparticle research. *Chem. Rev.*, 116:11128–11180, 2016.

[49] F. Spinozzi, C. Ferrero, M. G. Ortore, A. De Maria Antolinos, and P. Mariani. GENFIT: software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in-solution. *J. App. Cryst.*, 47:1132–1139, 2014.

[50] A.V. Hill. The possible effects of the aggregation of the moleculs of haemoglobin on its dissociation curves. *J. Physiol.*, 40:iv–vii, 1910.

[51] I. Langmiur. The adsorption of gases on pane surfaces of glass, mica and platinum. *J. Am. Chem. Soc.*, 40:1361–1403, 1918.

[52] J.P. Hansen and I.R. McDonald. The theory of simple liquids. *Academic Press:London*, 1986.

[53] H. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide protein data bank. *Nat. Struct. Biol.*, 10:980, 2003.

[54] M .Lundqvist M., J. Stigler, T. Cedervall, T. Berggard, M.B. Flanagan, G. Elia, and K. Dawson. The evolution of the protein corona around nanoparticles: A test study. *ACS Nano*, 5:7503–7509, 2011.

[55] C. Rocker, M. Potzl, F. Zhang, J. Wolfgang, J. Parak, and G.U. Nienhaus. A quantitative fluorescence study of protein monolayer formation on colloidal nanoparticles. *Nat. Nanotechnol.*, 4:577–580, 2009.

[56] R. Capamaccio, I. Ojea-Jimenez, P. Colpo, D. Gilliland, G. Ceccone, F. Rossi, and L. Calzolai. Determination of the structure and morphology of gold nanoparticle - hsa protein complexes. *Nanoscale*, 7:17653–17657, 2015.

[57] L.R. Barbosa, M.G. Ortore, F. Spinozzi, P. Mariani, S. Bernstorff, and R. Itri. The importance of protein-protein interactions on the ph-induced conformational changes of bovine serume albumin: A small-angle x-ray scattering study. *Biophys.J.*, 98:147–157, 2010.

[58] C. Minnelli, P. Moretti, G. Fulgenzi, P. Mariani, E. Laudadio, T. Armeni, R. Galeazzi, and G. Mobbili. A polaxamer-407 modified liposome encapsulating epigallocatechin-3-gallate in presence of magnesium:characterization and protective effect against oxidative damage. *International Journal of Pharmaceutics*, 552:225–234, 2018.

[59] L.F. Hernandez-Zimbron, R. Zamora-Alvarado, L. Ochoa-DelaPaz, R. Velez-Montoya, E. Zenteno, R. Gulias-Canizo, H. Quiroz-Mercado, and R. Gonzales-Salinas. Age-related macular degeneration:new paradigms for treatment and management of amd. *Oxid. Med. Cell. Longev*, 1:8374647, 2018.

[60] B. Frei and J.V. Higdon. Antioxidant activity of tea polyphenols in vivo: evidence from animal studies. *J. Nutr.*, 133(10):3275S–3284S, 2003.

[61] J.F Severino, B.A. Goodman, C.W. Kay, K.Stolze, D. Tunega, T.G. Reichenauer, and K.F. Pirker. Free radicals generated during oxidation of green tea polyphenols: electron paramagnetic resonance spectroscopy combined with density functional theory calculations. *Free Radic. Biol. Med*, 46(8):1076–1088, 2009.

[62] B. Zhang, R. Safa, D. Rusciano, and N.N. Osborne. Epigallogallate, an active ingredient from green tea, attenuates damaging influences to the retina caused by ischemia/reperfusion. *Brain Res.*, 1159:40–53, 2007.

[63] R. Galeazzi, P. Bruni, E. Crucianelli, E. Laudadio, M. Marini, L. Massaccesi, G. Mobbili, and M.Pisani. Liposome-based gene delivery systems containing a steroid derivative: computational and smalll angle x-ray diffraction study. *RCS Adv.*, 5:54070, 2015.

[64] E. Laudadio, C. Minnelli, A. Amici, G. Mobbili, and R. Galeazzi. Liposomal formulations for an efficient encapsulation of epigallocatechin-3-gallate: an in-silico/experimental approach. *Molecules*, 23(2):441, 2018.

[65] N. Mignet, J. Seguin, and G.G. Chabot. Bioavailability of polyphenol liposome: a challenge ahead. *Pharmaceutics*, 5:457–471, 2013.

[66] G. Wu, H.A. Kant, W. Chiu, and K.Y.C. Lee. Effects of bilayer phases on phospholipid-poloxamer interactions. *Soft Matter*, 5:1496–1503, 2009.

[67] G. Dumortier, J.L. Grossiord, F. Agnely, and J.C. Chaumeil. A review of poloxamer 407 pharmaceutical and pharmacological characteristics. *Pharm. Res.*, 23(12):2709–2728, 2006.

[68] F. Szoka Jr and D. Papahadjopoulos. Procedure for preparation of liposomes with large internal aqueous space and high capture by reverse-phase evaporation. *Proc. Natl. Acad. Sci. U.S.A.*, 75(9):4194–4198, 1978.

[69] W.E Price and J.C. Spitzer. The kinetics of extraction of individual flavanols and caffeine from a japanese green tea (sen cha uji tsuyu) as a function of temperature. *Food. Chem.*, 50:19–23, 1994.

[70] E. Esposito, P. Mariani, M. Drechsler, and R. Cortesi. Structural studies of lipid-based nanosystems for drug delivery: xray diffraction (xrd) and cryogenic transmission electron microscopy (cryo-tem). *Springer International Publishing*, pages 861–889, 2016.

[71] G. Pabst, M. Rappolt, H. Amenitsch, and P. Laggner. Structural information from multilamellar liposomes at full hydration:full q-range fitting with high quality x-ray data. *Phys. Rev. E:stat. Phys. Plasmas Fluids Relat. Interdiscip.*, 62(3PtB):4000–4009, 2000.

[72] G.M. Di Gregorio, P. Ferraris, and P. Mariani. Wetting properties of dioloeoyl-phosphatidyl-choline bilayers in the presence of trehalose: an x-ray diffraction study. *Chem. Phys. Lipids*, 163(6):601–606, 2010.

[73] H. Binder and O. Zschornig. The effect of metal cations on the phase behavior and hydration characteristics of phospholipid membranes. *Chem. Phys. Lipids*, 115(1-2): 39–61, 2002.

[74] L. Djekic, D. Krajisnik, M. Martinovic, D. Djordjevic, and M. Primorac. Characterization of gelation process and drug rlease profile of thermosensitive liquid lecithin/poloxamer 407 ased gels as carriers for percutaneous delivery of ibuprofen. *Int. J. Pharm.*, 490(1):180–189, 2015.