UNIVERSITÀ
POLITECNICA
DELLE MARCHE

DOCTORAL THESIS

Key SAXS methodologies to unravel protein complexes, protein stability and textile microfibers degradation

Supervisor:
Prof. Francesco SPINOZZI

PhD Student:
Astra PICCININI

Co-Tutor:
Prof. Anna LA TEANA
Prof. Heinz AMENITSCH

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

Ancona, 2021-2022

Declaration of Authorship

I, Astra Piccinini, declare that this thesis titled, *Key SAXS methodologies to unravel protein complexes, protein stability and textile microfibers degradation* 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

Key SAXS methodologies to unravel protein complexes, protein stability and textile microfibers degradation

by Astra Piccinini

My PhD project focuses on different applications of Small-Angle X-ray Scattering, shortened to SAXS, a non-conventional technique that allows getting information about nano-objects in solution in a dimensional range between 1 and 100 nm and can quantify the particles behaviour in response to changes of the physical-chemical conditions. Through SAXS, samples can be analysed in their natural environment and features like shape, size, and low-resolution internal structure are easily obtained. The first part of my PhD focused on the application of SAXS in Biology, in particular on two Archaea proteins derived from the thermophilic organism *Sulfolobus solfataricus*, which are found in solution as tetramers and as a series of oligomers. The proteins above mentioned are the Archaea Deoxyhypusine synthase (DHS) and the translation elongation factor aIF5A, respectively: since they are universally conserved in all the kingdoms of life and aIF5A is highly overexpressed in different eukaryotic cancer cells, their characterization would be of great help to understand their mechanism of interaction. In order to obtain high quality SAXS data, it is important to perform experiments in monodispersed systems, which is the reason why the optimization of a Size Exclusion Chromatography (SEC) apparatus coupled with SAXS is crucial. However, depending on the investigated system, there could be no need to separate the sample into different fractions by means of chromatography, and the sample can directly undergo X-rays exposition. The second investigation discussed in this thesis, which was done in collaboration with Lisbon University and the affiliated spin-off Extremochem, regards the SAXS investigation of the stabilization capacity of some modified sugars on two model proteins, in order to test the ability of those cosolvents to preserve the protein folded state and prevent the aggregation. To describe proteins and sugars interactions, a thermodynamic model was specifically developed and successfully adopted for the analysis of the whole set of SAXS data. Last but not least, it is important to underline that SAXS not only can be used to perform experiments on biological samples but also on other substances and materials thanks to its versatility, which is one of its key features. For this reason, and considering my background on Environmental Sciences, great attention was also paid to one topic that goes beyond Biology, but which is as critical as the others: the emerging pollutants. Among this wide group of materials, we have focused our attention on textile microfibers derived from the laundry of clothes, which are usually not trapped by the water treatment plants and reach the sea where different degradation processes take place. Several types of microfibers among the most globally used in the market were chosen and subsequentially irradiated from a solar simulator to reproduce one year of solar exposition. Then, a periodically sampling was done in order to obtain a degradation profile as a function of the solar simulator exposure

time, and the samples were investigated with SAXS.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Small-Angle X-ray Scattering is a non-conventional technique widely used to study the structure of materials with nanometric size dispersed in solution. When SAXS is applied to biological macromolecules, a particular advantage is that they don't need to be crystalized and can be studied in their physiological environment, allowing the extraction of information such as the folding and unfolding state of proteins, the hydration properties and conformational changes that occur in solution. However, the set of samples investigated not only include biological macromolecules but also polymers and nanocomposites. More in general, SAXS is particularly versatile and essential for a wide range of scientific fields, including Biology, Physics, Chemistry, Material Science and Environmental Sciences. By taking advantage from the adaptability and flexibility of SAXS, three different systems have been studied in this thesis:

- The structural features of the Archaea translation factor aIF5A;
- Stabilizing properties of modified sugars on model proteins;
- Degradation degree of textile micorfibers after a long term simulation of sun exposure.

1.1 Systems studied

1.1.1 The translation factor IF5A

1.1.1.1 IF5A

Protein synthesis is a key process in the cells of all kingdoms. The core of the protein synthesis is the process by which the genetic information, copied from DNA and embedded in the messenger RNA (mRNA), is translated into amino acids sequences that will set up polypeptide chains. The regulation of this translation process is controlled by multiple factors. Among them, the translation factors cooperate actively during the initiation, the elongation, and the termination phase of protein biosynthesis. One such factor that takes part in the translation process is the so-called translation elongation factor IF5A. In the past, IF5A was classified as a translation factor which mostly performs its activity during the initiation phase of the protein synthesis, related to the methionyl-puromycin synthesis, increasing also its probability in taking part in the first peptide bond [12] and in the starting codon selection [41]. Subsequently, important discoveries have been made and to date it is known that IF5A is involved in the elongation as well as in the termination phase of protein synthesis. The crucial role played by IF5A during the elongation phase of proteins is rescuing ribosomes from stalling in presence of particular polyproline stretches, which because of their conformation would otherwise stop the extension of the peptide chain. To perform this action, IF5A relies on its unique post-translational modification of one lysine residue into hypusine. By means of hypusine, IF5A is able to bind the trinucleotide cytosine-cytosine-adenine (CCA) 3' end of the P-region of the transfer RNAs

(P-tRNA), stabilize its geometry, and allow the formation of the peptide bond of the following amino acids. IF5A is a highly conserved protein with homologs in Archaea, Eukarya and Bacteria domains. In particular, in Eukarya and Archaea (such as the model organism *Sulfolobus solfataricus* used in this study) the lysine residue that undergoes hypusination is the residue 50 and the residue 36, respectively, whereas in Bacteria the hypusination pathway is not present and instead β-lysination takes place. In humans, a strong correlation between the overexpression of the hypusinated form of IF5A and several diseases, such as malaria, diabetes and cancer have been established. In particular, in Eukarya eIF5A is present in two isoforms: eIF5A1, which is ubiquitous and overexpressed in some types of cancers, and eIF5A2 that is rarely present in normal tissues but much more detectable in lots of malignancies [42]. For these reasons, the hypusination pathway of IF5A has become an important therapeutic target to be considered. To note, the IF5A activity, which is not only related to the protein synthesis but also to the RNA binding and RNA decay capacity, is possible thanks to its unique post-translational modification, which in Eukarya is catalyzed by two enzymes: Deoxyhypusine synthase (DHS) and Deoxyhypusine hydroxylase (DOHH). In Archaea the protein is hypusinated but only the first step of the hypusination pathway, catalyzed by a protein homologous to DHS, has been characterized, while the archaeal counterpart of DOHH has not been identified yet [59]. Several inhibitors have been tested in order to discourage the first step of the hypusination pathway, which is the interaction between IF5A and DHS, but their non-specificity resulted in several negative side effects that brought to the necessity of having a more detailed insight on the proteins structures and interactions.

1.1.1.2 Post-translational modification

Even though the eukaryotic hypusination pathway is well studied, little is known about the same mechanism in Archaea, although the structure of the N-terminal domain of IF5A, where the lysine residue subjected to hypusination is present, is very much conserved. For all these reasons, the first step of hypusination, which is the interaction between IF5A and DHS, is particularly interesting to study and, since the archaeal model is less complex then the eukaryotic one, it can be used as a model system to shed light on the hypusination process in superior organisms. Deoxyhypusine synthase, represented in Fig. 1.1, is the enzyme that catalyses the first step of the hypusination pathway of IF5A and this reaction, which takes place by means of the NAD dependent transfer of one group of spermidine, is extremely conserved. The same degree of conservation characterizes also DHS and IF5A, as already explained before. aDHS from the archaeal organism used in this PhD work is composed of four identical subunits of around 36 kDa (41 kDa in humans), with four active sites respectively organised in each unit. The active site of the monomer is not easily accessible since is hidden inside the monomeric structure in a small region that forms a tight tunnel at the end of which the interaction between substrate and enzyme takes place [19]. In Fig. 1.1, the recently crystallized structure of DHS from the archaeal organism *Pyrococcus horikoshii* is reported.

In Eukarya, eIF5A undergoes a post-translational modification in its Lysine residue (Lys 50) by means of two enzymatic steps. First of all, DHS transfers the 4 amino-butyl moiety of spermidine to the ϵ-amino group of one specific lysine residue and it forms

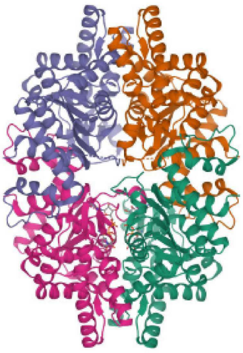


Figure 1.1: Structural representation of the archaeal enzyme Deoxyhypusine Synthase with the four subunits of the protein colored differently. The PDB derives from *Pyrococcus horikoshii*, PDB code 7CMC. [17]

the deoxyhypusine residue. Then, DOHH hydroxylates the intermediate and it forms the final hypusine residue leading to the mature form of eIF5A.

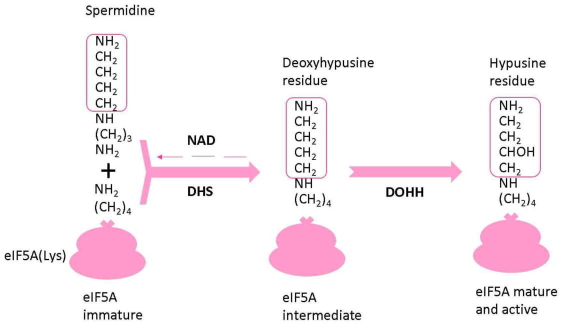


Figure 1.2: Hypusination pathway rearranged from [59]

1.1.1.3 Archaea IF5A structure

Considering the conservation degree between IF5A and DHS in the Eukarya and Archaea domains and the post-translational modification on IF5A, which is present in both aIF5A and eIF5A [11], the Archaea protein has been chosen as the most suitable for the analysis of the interaction between the protein and the enzyme. The model organism used was

Sulfolobus solfataricus, an Archaea belonging to Crenarchaeota. In particular *Sulfolobus solfataricus* is a thermophilic and acidophilic organism that grows in extreme conditions. Its optimal temperature is 80° C and the pH ranges between 2 and 4, although it is able to maintain an internal pH value near 6.5 [66]. aIF5A is a small protein of 15 kDa isolated for the first time from the *Sulfolobus acidocaldarius* [9]. This protein bears a resemblance to the eukaryotic factor concerning its structure and function. The crystal investigation of aIF5A derived from *Pyrococcus horikoshii* shows similarities between the two homologous proteins. Two domains characterize aIF5A: the N-terminal domain, which contains the lysine residue that undergoes hypusination, composed by six β-stranded anti-parallel β-sheets and one α-helix, and the C-terminal domain, with a oligonucleotide-binding fold with five stranded anti-parallel β sheets and two α helices [68]. The N-terminal domain holds the lysin residue (Lys 37 in *Pyrococcus horikoshii*), which is responsible for the post-translational modification of the protein with the formation of the hypusine residue [94]. In Eukarya, IF5A is a protein of 17 kDa composed of two domains: the N-terminal domain, formed by one α helix and several β-strands among which the hypusine residue is embedded, and the C-terminal domain, composed of one α helix, five β-strands and an oligonucleotide binding fold that might allow the protein to bind RNA [68].

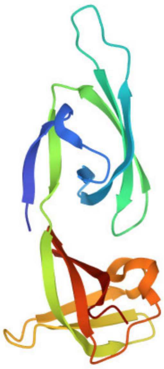


Figure 1.3: Translation factor aIF5A from *Pyrococcus horikoshii* [94], PDB code 1IZ6.

1.1.2 Modified sugars

1.1.2.1 Protein stabilization and stabilizing agents

Proteins are complex sequences of amino acids that arrange in different ways. Apart from the primary structure, which is characterized by the succession of amino acids united by peptide bonds, other three levels of structure are possible; starting from the secondary structure, which is a repetition of regular protein motifs able to build the macromolecular backbone, to the third and fourth structures, which represent the three-dimensional folding and the merge of more proteins chains, respectively. The tertiary structure is directly related to the protein's function and typically holds the least free energy content that corresponds to the maximum stabilization of the protein itself. One protein can maintain a certain conformation by means of weak interactions between itself and the surrounding environment that is usually occupied by water molecules. One of the main factors that regulates the protein folding is the hydrophobic effect. This interaction prevents the non-polar amino acids of the protein from coming into contact with water, forcing them to compartmentalize the solvent within hydrophilic amino acids. Hydrogen bonds also occur between hydrogen atoms covalently bound to an electronegative atom and another electronegative atom bearing one lone pair of electrons, such as oxygen or nitrogen. This interaction mostly stabilizes the secondary structure of proteins such as α-helices and β-strands, but it can also be found to stabilize their tertiary structure. Additionaly, the stability of the tertiary structure can be enhanced by disulfide bonds, which stabilize proteins by reducing the entropy of the unfolded state. They are formed by a single covalent bond between two sulphur atoms derived from two cysteines. The interactions between amino-acids residues with opposite charges gives rise to the attractive ionic interactions, also called salt bridges. Last but not least, weak forces, also known as van der Waals interactions, can be classified into three groups:

- Two permanent dipoles (Keesom);
- Permanent and induced dipoles (Debye);
- Instantaneous and induced dipoles (London).

In summary, the interactions between backbones usually stabilizes the secondary structure, whereas the interactions between side chains are more prone to participate in the stabilization of the tertiary and quaternary structure. All these forces are briefly described in table 1.1.

Despite the intra and intermolecular forces that take place to stabilize the folding state of one protein, several stresses may lead to the unfolding and the loss of some particular functions, such as high temperatures, pH changes, or mechanical stresses. Nowadays, proteins are increasingly used as therapeutics molecules in several diseases and, for this

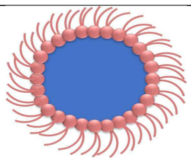
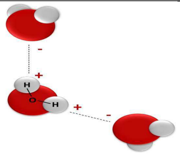
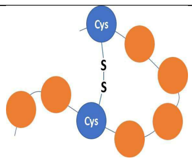
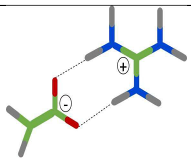
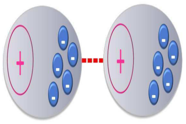
Hydrophobic effect	Hydrogen bonds	Disulfide bonds	Salt bridges	van Der Waals
				

Table 1.1: Stabilizing forces

reason, it became even more crucial solving problems related to their stability [70]. In order to maintain the right protein conformation, it is possible to rely on particular stabilizing agents, which can comprehend a multitude of molecules such as sugars, salts, amino acids, and other polymers able to increase the energy barrier between the folded and the unfolded state of proteins. Cosolvents, for example, interact with water molecules around the protein preserving its hydration shell and maintaining its stability. A summary of the pharmaceutical excipients used to stabilize proteins is reported in Table 1.2, where a series of compounds ranging from small to large molecules are shown. However, it is necessary to take into account that pharmaceutical excipients may be not completely stable or pure and can alter the stability profile of the protein itself [37]. In this thesis we focused our attention on five particular types of modified-sugars supposed to stabilize two model proteins against denaturation induced by increasing temperatures. To have a complete idea about the mechanism that happens in solution, both structural and thermodynamic features of the proteins have been considered. Results of this work have been published in Life journal [63].

Category	**Representative examples**	**General comments**
Buffering agents	Citrate, acetate histidine, phosphate,Tris	-Maintain solution pH -Buffer-ion specific interaction with proteins
Amino acids	Histidine, arginine, glycine, proline, lysine, methionine	-Specific interactions with protein -Antioxidants -Buffering and tonicifying agents
Osmolytes	Sucrose, trehalose, sorbitol, glycine, proline, glutamate, glycerol, urea	Natural compounds that stabilize proteins and macromolecules against environmental stress (temperature, dehydration)
Sugars and carbohydrates	Surcrose, trehalose, sorbitol, mannitol, glucose, lactose	Protein stabilizer in liquid and lyophilized states
Proteins and polymers	HSA, gelatin, PVP, PLGA, PEG	-Competitive inhibitor of protein adsorption -Lyophilization bulking agents -Drug delivery vehicles
Salts	Sodium chloride, potassium chloride, sodium sulfate	-Tonicifying agents -Stabilizing or destabilizing effects on proteins with anions
Surfactants	Polysorbate 20 and 80	-Competitive inhibitor of protein adsorption -Competitive inhibitor of protein surface denaturation -Liposomes as drug delivery vehicles
Chelators and anti-oxidants	EDTA, DTPA, amino acids, ethanol	-Bind metal ions -Free radical scavengers
Preservatives	Benzyl alcohol, m-cresol, phenol	Prevents microbial growth
Specific ligands	Metals, ligands, amino acids, polyanions	-Bind proteins and stabilize native conformation against stress

Table 1.2: Summary of stabilizing agents and their properties rearranged from [37]

1.1.3 Textile microfibers and long term solar simulation

1.1.3.1 Plastic pollutants in marine environment and photoaging

The aquatic environment is one of the most threatened worldwide. It hosts a lot of waste from the top to the bottom of the water column, which sometimes is not visible but prone to threaten ecosystems. The increasing anthropogenic pressure and the consumerism that leads the society nowadays is bringing to an unhealthy overuse of some materials, with high demand in the global market. Among them plastic stands out. Million tonnes of plastic are produced annually and most of them are discarded in the environment ending up in the sea. Once in the sea, plastic debris are ingested by animals and stored in their bodies. Their dangerousness is not only related to the debris itself, but also to the pollutants and the additives that are present on their surfaces and whose accumulation may bring serious consequences to the health of the victims, becoming even more dangerous with the biomagnification process at higher food-chain levels. The smaller is the debris, the more it is likely that it won't be trapped by the water treatment plants and it reaches the sea, where several degradation processes take place, shrinking it enough for entering the bloodstream of animals. The main degradation processes might be divided into three categories: (i) biodegradation exerted by microorganisms; (ii) weathering developed by waves motion/abrasion with the bottom of the sea; (iii) photodegradation performed by the solar light. Among the latter category, photoaging is a process widely overlooked, even though it was shown that UV irradiation brings to the formation of several free radicals species on the plastic surface [97]. Free radicals form between the C-C and C-H bond and bring to the generation of a radical carbons, which react with other molecules of the chain polymer, including oxygen, with the development of peroxyl radicals, which will restart the radical generation cycle. All these reactions result in continuous chain scissions and linkages that make the polymer chain unstable and prone to incur breaks. The degradation of plastic debris into smaller pieces leads to the formation of microplastics and nanoplastics (with at least one dimension in the order of 10^{-6} m and 10^{-9} m, respectively). According to a molecular dynamic study performed by Hollóczki and Gehrke [35], nanoplastics are able to modify the secondary structure of proteins with a consequent change in their function, which sometimes leads to their denaturation.

1.1.3.2 Textile microfibers as emerging pollutants

A big source of microplastics has recently been discovered to origin from textile fibers. Each year millions of tons of fibers are produced worldwide, whose 60% are of synthetic origin [8]. One single garment during the laundering of clothes may release more than 1900 fibers per wash that are not stopped by any filters and directly enter into the aquatic environment [15]. Textile fibers are not usually taken into account during waste collection, although they are among the most common pollutants found in the marine environment and this underestimation makes them even more harmful because overlooked. A recent study, developed by Polytechnic University of Marche (Ancona, Italy), provides a full-scaled characterization of microplastics ingested from around 500 marine organisms in the Adriatic sea [4]. Results of the study revealed that the number of microfibers (derived from the degradation of textile fibers) found in the investigated marine organism was higher than that of microplastics. Additionally and even more unpredictable, the chemical characterization performed with micro-Fourier transform infrared spectroscopy (μFTIR) showed a predominance of microfibers coming from natural polymers compared to that coming from synthetic polymers. This mismatch between the synthetic global fibers production and the natural waste collection, mainly cellulose-based, suggests a potential difference in their degradation pathway. For this reason, in this PhD study four textile microfibers were chosen, two of natural origin (linen and cotton) and two of synthetic origin (polyester and polyamide,better known as nylon). Moreover, another ubiquitous fiber commonly present as a marine environmental pollutant was considered: cellulose acetate from cigarette butts.

Chapter 2

Materials and Methods

2.1 X-rays

2.1.1 Introduction to X-rays

In 1895 Whilem Conrad Röntgen discovered X-rays while he was carrying on experiments on vacuum tubes. Since then, X-ray have been widely used in different fields, first of all in medical diagnostics devices. X-rays are electro-magnetic radiations with a dual wave-particle nature, as they simultaneously behaves as corpuscles (called photons) and as waves (Fig. 2.1). The X-ray wavelength λ is comprised between 0.01 and 10 nm, and the energy carried by each photon, according to Planck's law, $E = hc/\lambda$ (where h is the Planck constant and c the light velocity), ranges from 0.1 to 100 keV [90].

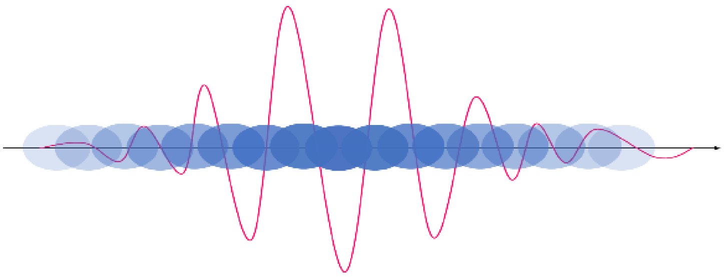


Figure 2.1: Wave-particle duality representation.

X-rays are usually generated in laboratory by means of cathodic tubes where, through the thermionic emission, electrons are released from a cathode and accelerated towards an anode. When electrons hit the anode, two types of radiations are produced: one depends on the electrons' deceleration effect, called Bremsstrahlung, leading to the development of a wide spectrum of wavelengths, and the other one is based on the type of the target material, which shows peaks of intensity centered in characteristic values of λ. In detail, when the electron passes through an atom, its direction is deflected by the electromagnetic field nearby the atomic nucleus. Because of the energy conservation principle, the kinetic energy lost by the electron during its deflection is emitted as Bremsstrahlung radiations (Fig. 2.2), a continuous radiation distribution emitted during this deceleration process.

When a high energy electron impacts on an atom, an inner electron could be ejected and a valence hole is formed, making the atom unstable. Afterwards, this gap is filled through an electron coming from an outer shell, leading to an emission of energy in the form of characteristics X-rays, whose wavelength depend on that specific target material (Fig. 2.3).

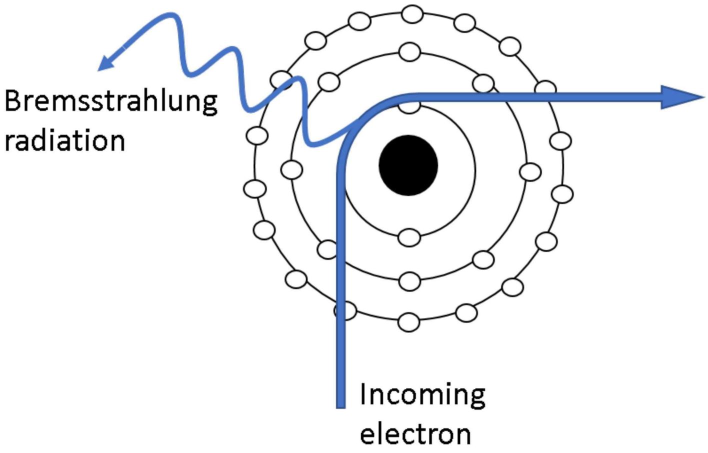


FIGURE 2.2: Bremsstrahlung principle.

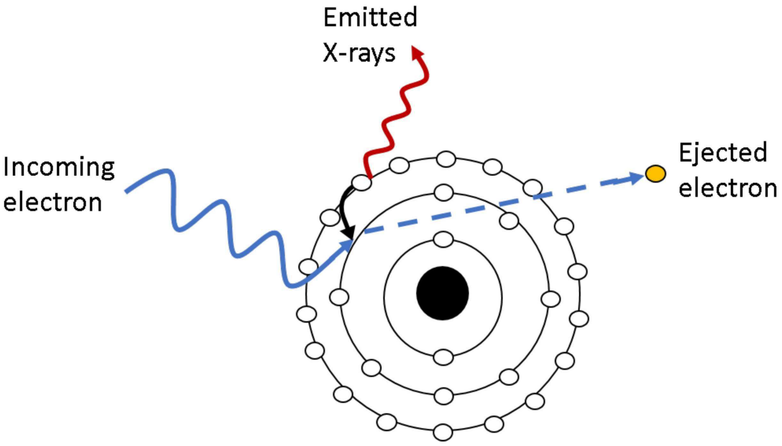


FIGURE 2.3: Illustration of the emission of characteristic X-rays.

2.1.2 Synchrotron

Synchrotrons are large scale facilities able to provide a powerful and continuous source of electromagnetic waves, including X-rays, of a wide range of wavelengths. Charged particles such as electrons are accelerated in a booster and then sent to a bigger circular path where they are forced to deflect by magnetic fields (Fig. 2.4). Each change in direction brings to a loss of energy that is emitted in the form of the so-called synchrotron radiation. The electrons' deceleration process produces Bremsstrahlung and this results in a continuous wavelength spectrum. The synchrotron light is then channelled into tangent lines called beamlines, where the exact wavelength will be selected and used to irradiate a sample. One of the synchrotrons' advantages is the extremely powerful light that allows scientists to extract information much more detailed in respect to the one that is possible to achieve by the common laboratory X-ray production. On one hand, thanks to this powerful radiation, samples don't need to be highly concentrated and the time required for the analysis shortens a lot. On the other hand that strong light may damage the samples, which will not be reusable anymore.

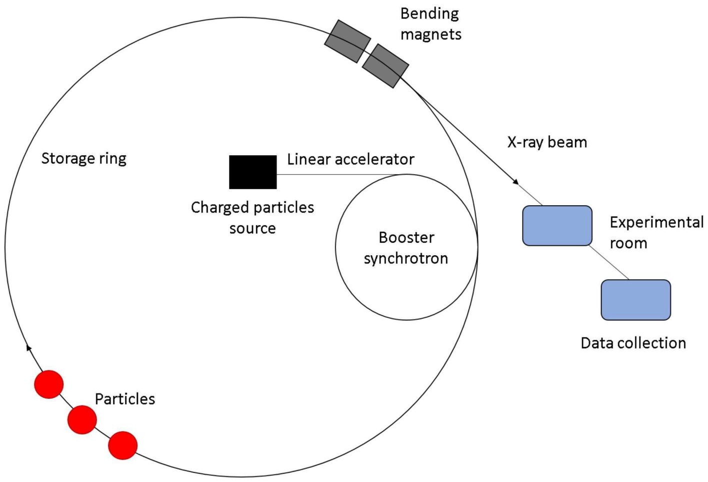


Figure 2.4: Schematic synchrotron representation.

2.1.3 Small angle X-ray scattering

A high number of elastic scattering techniques based on X-ray are commonly used in several fields of science, and each of them has a different resolution that make them suitable to analyse the structure of samples with different degrees of order. The most commonly used X-rays elastic scattering techniques are: X-ray crystallography, X-ray diffraction and small-angle X-ray scattering shortened to SAXS (Fig. 2.5). SAXS is one of the most suitable X-ray tool to provide structural information on samples containing nanosized particles randomly dispersed in solution [24]. Since biological macromolecules, such as proteins, or assemblies, such as lipid vesicles, have dimensions that fall in the nanometric regime, SAXS has been successfully applied in Biology. As far as proteins are concerned, folding and unfolding states, conformational changes, interactions between molecules in solution, hydration properties and quaternary structures are some of the many protein properties that can be investigated with SAXS [88, 65, 79, 45]. Although SAXS does not allow getting information about the atomic positions of the macromolecule, which is a key feature of crystallography, it is able to measure the isotropic scattering pattern of all the particles in solution, which contains information about particle size and shape as well as particle-particle interactions [76]. In addition to biological materials, SAXS can be applied to a wide range of materials, such as polymers, nanocomposites, metallic alloys, among many others. The typical SAXS investigated dimensions range between 1 to 100 nm, but they can be extended to larger or smaller dimensions by using ultra-small-angle X-ray scattering (USAXS) and wide-angle X-ray scattering (WAXS), respectively. The huge advantage of SAXS in respect to other elastic scattering techniques based on X-rays is that samples don't need to be crystallized and can be investigated in their natural environment, which is crucial when macromolecules like proteins are studied in order to identify their biological activity in solution.

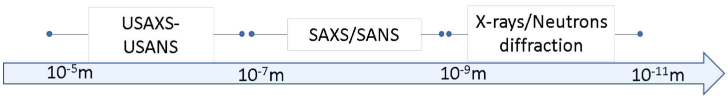


FIGURE 2.5: SAS resolution techniques.

2.1.3.1 SAXS theory

Before the polychromatic X-ray beam generated by a synchrotron source hits the sample, a series of operations are done in order to get an incoming beam with the proper features. The right wavelength is selected through a monochromator, typically between 1 and 3 Å, and a collimation system narrows the beam using slits and pinholes, in order to well separate the beam scattered by the sample, with wave vector $\mathbf{k} = (2\pi/\lambda)\hat{\mathbf{s}}$, from the direct beam, with wave vector $\mathbf{k}_0 = (2\pi/\lambda)\hat{\mathbf{s}}_0$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}_0$ being unit vectors associated to the directions of the two beams. Notice that the wave vectors have dimensions of the reciprocal of a length and are typically expressed in Å^{-1}. Accordingly, their space is referred to as the *reciprocal space*. When the sample is hit, part of the radiation is transmitted unaffected and blocked by the beam stop placed in the detector, and the other part, scattered in many directions, is measured by the different cells of the detector (Fig. 2.6). The SAXS scattering patterns are collected and expressed as the intensity of X-rays as a function of the scattering vector $\mathbf{q}$, which is defined as $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$. To note, since for elastic scattering processes the two wave vectors have the same modulus, the modulus q of the scattering vector only depends on the wavelength and on the angle 2θ formed by the direct and the scattering beam directions ($\hat{\mathbf{s}}_0$ and $\hat{\mathbf{s}}$, respectively), according to $q = (4\pi/\lambda)\sin\theta$. When particles have not a preferential orientation so that they are completely randomly oriented, the scattering pattern is isotropic: all the detector's cells placed at the same distance from the direct beam position (the same angle 2θ), on average, will measure the same intensity. To note, the sum of all such intensities if a typical operation applied to raw SAXS data referred to as *radial average* aimed to improve the statistics of the detector counts. Hence, for randomly oriented particles the SAXS intensity only depends on the modulus q of the scattering vector. The dimensional ranges that can be investigated with SAXS are directly related to the q-range and its reciprocal relation with real space. Low q values (corresponding to low 2θ angles) hold information of big particles, whereas high q have features of the smallest ones. There is a proper q range within it is possible to obtain details of the sample under anlysis and it ranges between $q_{\min}$ and $q_{\max}$. Indeed, considering D a typical particle size, the optimum q-range can be calculated as follow:

$$\begin{aligned} D_{\min} &\sim \frac{\pi}{q_{\max}} \\ D_{\max} &\sim \frac{\pi}{q_{\min}} \end{aligned} \tag{2.1}$$

The allowed values of $q_{\min}$ and $q_{\max}$ are strictly dependent on the sample-to-detector distance and on the collimation system. In detail $q_{\min}$ is related to the collimation because if the direct beam is not narrowed enough it can interfere with the scattering pattern of the biggest particles covering their signal. $q_{\max}$ on the other hand, also depends on the reduction of the scattering intensity and the increase of background noises. According to

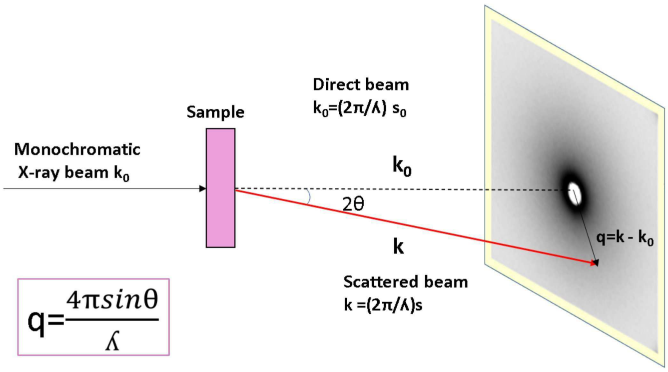


Figure 2.6: Schematic SAXS experiment.

the elastic Thomson scattering process of electromagnetic waves, which is the predominant scattering effect related to the interaction of X-rays with matter, when an atom is hit by X-rays, its electrons start oscillating at the same frequency (and wavelength) of the incident beam, creating spherical waves.

In a SAXS experiment, the scattering intensity of the particles is measured by the difference between the signal of the sample (containing particles in solution) and the signal of the solvent (both corrected by the detector efficiency and by the corresponding transmissions as well as by the signal of the empty cell and by the one of the natural or electronic background). Notice that, in order to achieve this result, *four* independent SAXS patterns and transmissions should be measured, corresponding to sample, solvent, empty cell and background. The scattering intensity of the particles, more precisely referred to as the *macroscopic differential coherent scattering cross-section* and indicated by $\frac{d\Sigma}{d\Omega}(q)$, is related to the electron density of the sample according to the fundamental or *master* equation,

$$\frac{d\Sigma}{d\Omega}(q) = \frac{1}{V}\left\langle \left| \int_V \delta\rho(\mathbf{r})\, e^{i\mathbf{q}\cdot\mathbf{r}}\, dV \right|^2 \right\rangle . \tag{2.2}$$

In this equation, V is the volume of the sample irradiated by X-ray beam, $\mathbf{r}$ is the position vector and the angular brackets represent the average over the possible states (i.e. oligomers, disordered conformers, among others), orientations and positions of the particles in the sample during the measuring time. The function $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_0$ is the local excess scattering length density (SLD), which corresponds to the difference between the electron density of the samples in the infinitesimal volume element dV around the position $\mathbf{r}$ and the constant electron density of the solvent, both multiplied by the classical radius of the electron ($r_e = 0.28 \cdot 10^{-12}$ cm). From a mathematical point of view, the integral in Eq. 2.2 represents the *Fourier transform* of the function $\delta\rho(\mathbf{r})$. The normalized amplitude of the X-ray wave scattered by one particle can be defined by the Fourier transform of the particle scattering length density, according to

$$F(\mathbf{q}) = \frac{1}{f} \int_{V_p} \delta\rho(\mathbf{r})\, e^{i\mathbf{q}\cdot\mathbf{r}}\, dV \tag{2.3}$$

where the integral only spans over the volume V_p of the particle and f is the zero-angle scattering amplitude,

$$f = \int_{V_p} \delta\rho(\mathbf{r}) dV \tag{2.4}$$

From a dimensional point of view, f is a length. The typical unit used in SAXS is 10^{-12} cm. We notice that f corresponds to the number of *excess* electrons (multiplied by r_e), i.e. the difference between the number of electrons provided by all the atoms of the particles and the number of electrons of the solvent molecules that occupies the same volume V_p of the particles, this latter equal to $\rho_0 V_p$. It can be shown, that when the system if formed by identical and well separated particles (dilute systems, see Fig. 2.7), the SAXS intensity is given by

$$\frac{d\Sigma}{d\Omega}(q) = nP(q) \tag{2.5}$$

where n is the number density of the particles in solution (simply related to their molar concentration C by $n = CN_A$, N_A being Avogadro's number). The term $P(q)$ is the so-called particle *form factor*, defined by $P(q) = f^2 \hat{P}(q)$, where $\hat{P}(q)$ (defined as the *normalized form factor*) is the orientational average (i.e. the average over the polar angles $\alpha_{\mathbf{q}}$ and $\beta_{\mathbf{q}}$ of the vector $\mathbf{q}$) of the square modulus of the normalized particle amplitude, according to $\hat{P}(q) =< |F(\mathbf{q})|^2 >_{\alpha_{\mathbf{q}},\beta_{\mathbf{q}}}= \int |F(\mathbf{q})|^2 \sin\beta_{\mathbf{q}} d\alpha_{\mathbf{q}} d\beta_{\mathbf{q}}$. Notice that

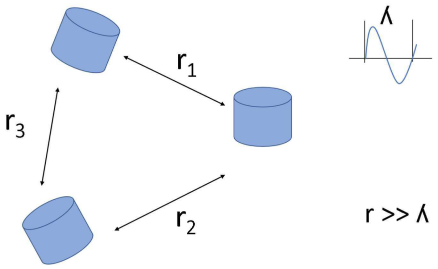


FIGURE 2.7: System are considered diluted when distances between particles are bigger than the wavelength.
[76].

the dimensions of $P(q)$ are a square of a length. The typical unit used in SAXS is 10^{-24} cm^2, that corresponds to 1 barn. When particles are not only identical but also *homogeneous*, a situation that occurs when the electron density within the particle volume is constant and different from the one of the solvent, $\frac{d\Sigma}{d\Omega}(q)$ can be written as follows:

$$\frac{d\Sigma}{d\Omega}(q) = n(\Delta\rho)^2 V_p^2 \hat{P}(q) \tag{2.6}$$

where $\Delta\rho = \rho_p - \rho_0$ is the so-called *contrast* of SLD between particle and solvent. In Fig. 2.8 the normalized form factors of homogeneous particles with a simple geometrical shape (a sphere, a cylinder and an ellipsoid) are shown. When the system is formed by particles with different size or shape (polydisperse samples), the *average form factor* of the system is

$$P(q) = \frac{1}{n}\sum_{j=1}^{m} n_j P_j(q) = \sum_{j=1}^{m} X_j P_j(q), \tag{2.7}$$

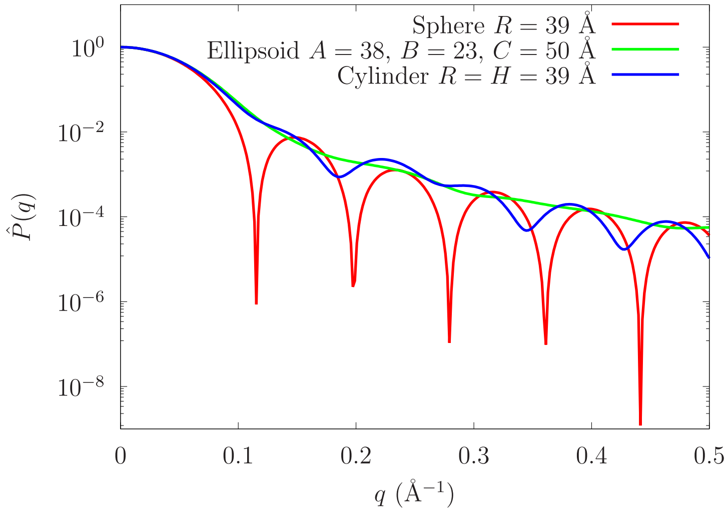


Figure 2.8: Simulated normalized form factors of a sphere, an ellipsoid and a cylinder [7].

where the sum runs over all the m kinds of particles in solution, each one with its number density n_j and its form factor $P_j(q)$, and $n = \sum_{j=1}^{m} n_j$ is the total particle number density. To note, we have here introduced the mole (or molecular) fraction of the j-particle, $X_j = n_j/n$. More in general, for polydispersed system of interacting particles, the scattering intensity $\frac{d\Sigma}{d\Omega}(q)$ results

$$\frac{d\Sigma}{d\Omega}(q) = n\, P(q) S_M(q) \tag{2.8}$$

In Eq. 2.8 the term $S_M(q)$ is the *effective structure factor* (or *measured structure factor*), which depends in a complex manner, here not shown, on $P_j(q)$ and on the orientational average of the normalized scattering amplitudes of the particles, $\hat{P}_j^{(1)}(q) =< F_j(\mathbf{q}) >_{\alpha_{\mathbf{q}},\beta_{\mathbf{q}}}$, as well as on the interactions among all the possible kinds of pairs of particles in solution. For example, if the system contains monomers and dimers of a protein, $S_M(q)$ depends on the monomer-monomer, monomer-dimer and dimer-dimer interactions. For homogeneous particles, Eq. 2.8 transforms to

$$\frac{d\Sigma}{d\Omega}(q) = n \; < (\Delta\rho)^2 > < V_p^2 > \; \hat{P}(q) S_M(q) \tag{2.9}$$

where we have introduced: the average normalized form factor,

$$\hat{P}(q) = \frac{\sum_{j=1}^{m} X_j V_{p,j}^2 (\Delta\rho_j)^2 \hat{P}_j(q)}{\sum_{j=1}^{m} X_j V_{p,j}^2 (\Delta\rho_j)^2}; \tag{2.10}$$

the average contrast,

$$< (\Delta\rho)^2 >= \frac{\sum_{j=1}^{m} X_j V_{p,j}^2 (\Delta\rho_j)^2}{< V_p^2 >}; \tag{2.11}$$

and the average squared volume of the particles,

$$< V_p^2 >= \sum_{j=1}^{m} X_j V_{p,j}^2. \tag{2.12}$$

In the absence of interactions among particles, a situation that formally occurs at infinite dilution, $S_M(q) = 1$ and the behaviour of $\frac{d\Sigma}{d\Omega}(q)$ is only due to $P(q)$. In these conditions, $\frac{d\Sigma}{d\Omega}(q)$, at low q, is described by a Gaussian centered at $q = 0$ (see curves in Fig. 2.8). On the other hand, when particles interact an upward or a downward deviations in respect to the ideal Gaussian behavior can be found (see the red curve in Fig. 2.9) Upward deviations, leading to a positive concavity at low q (green curve in Fig. 2.9), underline an aggregation behaviour of particles in solution, whereas downward deviations, that may determine the presence of a peak in the low q region (blue curve in Fig. 2.9), are representative of repulsive effects.

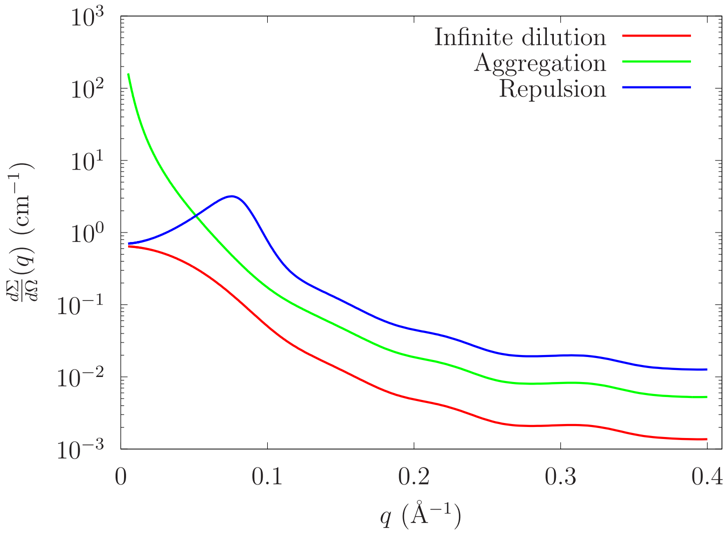


Figure 2.9: Trend of three SAXS curves where the low q profile changes according to the interaction of the particles in solution, which is accounted for by the structure factor.

A model independent shape analysis can be achieved by means of the Kratky representation, where data are plotted with $q^2\frac{d\Sigma}{d\Omega}(q)$ as a function of q. This plot is able to identify the conformation of proteins in solution with a simple visual inspection. The folding or unfolding state of a protein is evident by looking at the curve. In fact, a broad peak underlies the compactness of the protein, with the position dependent on the protein size, whereas the plateau-like trend is typical of a protein loss of the native structure resulting in an unfolding process.

To get information about the average size of the particles under analysis, it is necessary to rely on one particular approximation called Guiner approximation. According to the Guiner law (Eq. 2.13), for dilute systems at small q the curvature of the Gaussian function that approximates $\frac{d\Sigma}{d\Omega}(q)$ is only determined by the overall particle size.

$$\frac{d\Sigma}{d\Omega}(q) = \frac{d\Sigma}{d\Omega}(0)\exp\left(-\frac{q^2R_g^2}{3}\right) \tag{2.13}$$

As a matter of fact, by means of the Guiner law, the radius of gyration R_g, which by definition corresponds to the root-mean square of the distances from the mass centre of the scattering electrons, can be estimated.

2.2 SEC-SAXS

2.2.1 Chromatography

Chromatography means “colour writing”, a terms that refers to the first observations, which were accomplished to separate dyes by the German chemist Runge in the 1880s and to investigate plant pigments by Mikhail Tsvet in the 1900s. Over time, this technique has been refined and now it includes several methods aimed to separate all components of the sample under analysis. The main characteristics of this technique are the stationary phase, which firstly interferes with the sample, and the mobile phase that passes through the former phase and the sample itself. In Biology, chromatography can be divided into three main categories:

- Affinity chromatography, which separates components considering their binding affinity to one molecule or another;
- Ion exchange, where the charge of the molecule has a key role in the separation;
- Size Exclusion Chromatography that isolates molecules based on their size and shape [16].

2.2.2 Size Exclusion Chromatography

Size Exclusion Chromatography (SEC), also known as gel filtration, is characterized by a stationary phase consisting of beads with pores of different dimensions packed into a column. Beads are chemically and physically stable so that they are not prone to interact with the particles: their essential task is to retain the analyte compounds according to their molecular size. Bigger molecules are not able to enter the pores and thus they migrate through the space between them up to the end of the stationary phase : in fact, the bigger the molecule, the faster it exits the column ending up in the void volume V_0, which is the elution volume of particles' solution that do not interact with the column. The role of the mobile phase is the elution of the injected sample from the column. The chromatographic system can be connected to several types of detectors to measure the eluate characteristics, such as UV-VIS absorption, light scattering, viscosity, and refractive index. The time that the analyte spends from the injection to the detector is called *retention time*, and through the flowrate (volumetric flow, typically expressed in mL/min) it is directly related to the particle retention volume.

The goal of Size Exclusion Chromatography is to reach the maximum particle size resolution, which results in well-resolved and sharp peaks. This condition is generally achieved using a low flowrate, that allows molecules to have enough time to enter the pores. As a general rule, the best separation is obtained by using a low flow rate coupled with a long column. The SEC technique can be used to purify proteins from other proteins or aggregates, and it is also useful to extract information about their molecular weight (M_w) and size. In order to do so, the column must be calibrated with a set of molecules of

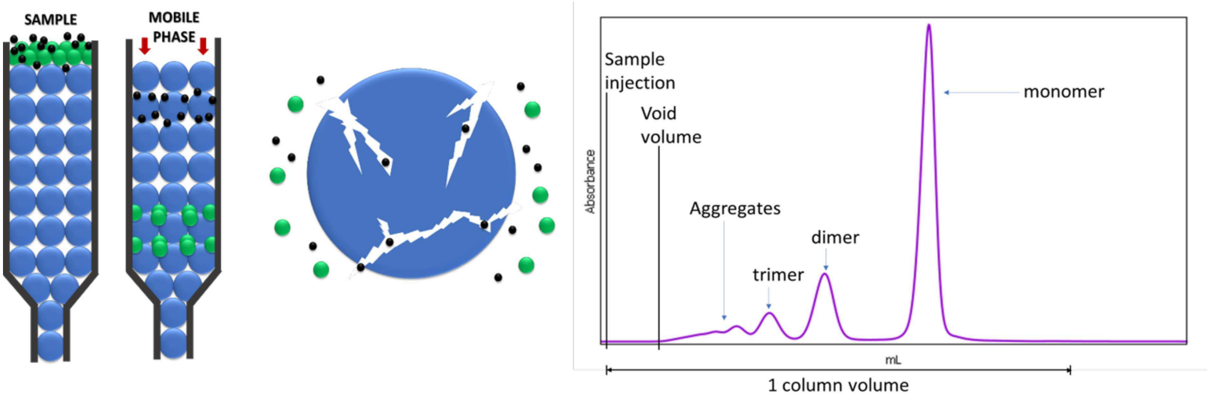


FIGURE 2.10: On the right a basic illustration of Size Exclusion Chromatography with a focus on the column beads that separate the sample on the basis of its dimension is represented. On the left a typical BSA chromatogram showing several peaks, which are characteristic of different protein aggregation states.

known molar mass that will allow developing a calibration curve from which it is possible to deduce the molecular weights of all the components of the sample. The calibration curve is useful whenever the sample taken as reference has similar size and shape of the sample under analysis, and this is due to the influence of the structure on the elution profile. However, another way to extract information about the absolute molar mass without using the standard calibration is through Multi-Angle Light Scattering.

2.2.3 Light Scattering

Light Scattering (LS) is a laboratory technique to measure the size of particles in dilute solutions with dimensions ranging from ≈ 10 nm up to ≈ 1000 nm. Two types of light scattering exist: static and dynamic. Static Light Scattering, also called Multi-Angle Light Scattering (MALS), measures the intensity of the scattered light at different angles and it is helpful to determine the molecular weight of the components of the mixture. Dynamic Light Scattering (DLS) or quasi-Elastic Light Scattering (qELS) analyses the fluctuations intensity of the scattered light, which are due to the Brownian motion of the particles in solution, and therefore to their dimensions [2]. In MALS experiments, a collimated light beam irradiates a sample that scatters light at different angles to several detectors. Each detector receives a signal proportional to the intensity of the light scattered from which the absolute molar mass is calculated. The main relationship that governs the MALS technique is the Zimm equation,

$$\frac{Kc}{R(2\theta)} = \frac{1}{M_w \hat{P}(q)} + 2A_2c \tag{2.14}$$

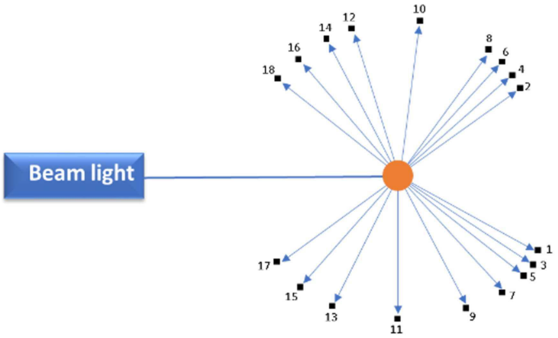
where $R(2\theta)$, which stands for Rayleigh ratio, corresponds to the ratio between the scattered intensity of the light at an angle 2θ and the intensity of the incident light ; c is the w/v sample concentration commonly expressed in mg/mL; M_w is the averaged molecular weight and A_2 is the second virial coefficient. The term K is the optical constant corresponding to

$$K = \frac{4\pi^2 n_0^2 \left(\frac{dn}{dc}\right)^2}{\lambda^4 N_A} \tag{2.15}$$

where n_0 is the solvent refractive index; dn/dc is the refractive index increment, which corresponds to the change of the refractive index of the solution with the change of the analyte concentration ; usually for proteins at room temperature and in aqueous solvents, it is considered constant and equal to 0.185 mL/g [1]. N_A stands for Avogadro's number, and λ is the wavelength of the scattered light. Finally, $\hat{P}(q)$ corresponds to the normalized form factor that depends on the average radius of gyration of the particles, according to the Zimm approximation,

$$\frac{1}{\hat{P}(q)} \approx 1 + \frac{q^2 R_g^2}{3} \tag{2.16}$$

where $q = 4\pi n_0 \sin\theta/\lambda$. This approximation holds when the condition $qR_g < 1$ is satisfied. However, if R_g is lower than $\approx \lambda/(4\pi n_0)$, such in the case of proteins, the form factor can be neglected ($\hat{P}(q) \rightarrow 1$). We notice that an extrapolation of the Zimm equation (Eq. 2.14) to both zero angle and zero concentration of $R(2\theta)$ measurements provides the molecular weight and second virial coefficient of the particles. In the case of experiments at a unique concentration, such as in the case for SEC-MALS approaches, only the molecular weight can be determined. The advantage of analysing samples with a wider range of angles is that the issue of the background noise at low angles is overcome. The MALS instrument used for the analysis in this PhD thesis is equipped with a multiangle detector which includes 18 angles, whose basic scheme is illustrated below.



Channel	Fixed angle	Channel	Fixed Angle
1	22.5°	10	81.0°
2	28.0°	11	90.0°
3	32.0°	12	99.0°
4	38.0°	13	108.0°
5	44.0°	14	117.0°
6	50.0°	15	126.0°
7	57.0°	16	134.0°
8	64.0°	17	141.0°
9	72.0°	18	147.0°

FIGURE 2.11: Multiangle Light Scattering used at the Elettra synchrotron during the SEC-SAXS analysis.

Chapter 3

Results

3.1 SEC-SAXS studies of the translation factor aIF5A

3.1.1 Recombinant N-term his tag aIF5A purification

The recombinant protein aIF5A derived from the thermophilic organism *Sulfolobus solfataricus* and expressed in *E. coli*, was purified using a standard protocol [10]. The purification process started with the growth of *E. coli* cells where the gene encoding for aIF5A has already been inserted. The plasmid used was pMCSG7 and the chosen strain for *E. coli* was Rosetta. Clones grew at 37° C in Lysogeny Broth medium rich in nutrients, where chloramphenicol and ampicillin were previously added to prevent the unpleasant growth of other bacteria apart from the selected clones, which already owned the gene for the antibiotic resistance. When the strains reached an optical density at 600 nm comprised between 0.4 and 0.8, they were induced by 0.5 mM isopropyl-β-D-1-tiogalattopiranoside

(IPTG), which is able to mimic the lactose metabolite that activates the LAC promotor and leads to a big amount of aIF5A synthesis. After 4 hours induction, cells were collected by centrifugation at 4° C in order to decrease some possible protease activities and then resuspended in lysis buffer (NaH_2PO_4 50 mM, NaCl 150 mM, lysozyme 1 mg/mL, phenylmethylsulfonyl fluoride (PMSF) 1 mM and imidazole 15 mM) for 30 minutes in ice. In addition to this chemical lysis, physical lysis was also applied. Cells were sonicated and centrifuged at 10000 rpm for 30 minutes at 4° C. Then, the lysate was subjected to purification by affinity chromatography using Ni-NTA agarose resin with a binding capacity of 50 mg/mL, which was previously equilibrated with the equilibration buffer (NaH_2PO_4 50 mM, NaCl 150 mM, imidazole 20 mM). The column beads were washed using 50 mL of washing buffer (NaH_2PO_4 50 mM, NaCl 150 mM, imidazole 20 mM) to eliminate all the impurities contained in the lysate and the protein was recovered in 5 mL of elution buffer (NaH_2PO_4 50 mM, NaCl 150 mM, imidazole 250 mM) with a higher concentration of imidazole, required to make a bond with the resin and take the place of the histidine tags. Then the protein was dialysed in dialysis buffer (Tris HCl pH 7.7-8 50 mM, KCl 150 mM) and it underwent protease cleavage to remove the histidine tags. Fig. 3.1 shows the protein purification in SDS-PAGE before the cutting of the histidine tail.

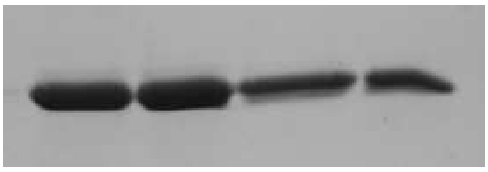


Figure 3.1: Detection of recombinant aIF5A his-tag protein in SDS-PAGE after the purification performed with affinity chromatography with nickel resin. The four bands represent the elutions of the protein recovered separately and then pulled together.

3.1.2 Recombinant Tobacco Etch Virus (TEV) protease purification

TEV is a small protein of about 27 kDa, and it is one of the most widely used proteases to remove histidine tags from other proteins. The purification of TEV (whose SDS-PAGE is reported in Fig. 3.2) in pMHTA238 in *E. coli* BL21 was done according to Blommel and Fox [14]. Cells grew over night at 37° C in a Lysogeny Broth medium modified with 0.8% glycerol, 0.375% aspartic acid pH 7, kanamycin and chloramphenicol and then the autoinduction was accomplished by means of a Terrific Broth consisting of 0.8% glycerol, 2 mM $MgSO_4$, 0.375% aspartic acid pH 7, 0.015% glucose, 0.5% α-lactose, plus the antibiotics previously mentioned. After 24 hours of autoinduction at 25° C, cells were collected via centrifugation and resuspended in lysis buffer (Hepes 50 mM pH 7.8, NaCl 200 mM, $MgCl_2$ 1 mM, Tris (2-carboxyethyl) phosphine (TCEP) 0.3 mM, Benzonase 4 U/mL) before sonication and centrifugation at 15000 rpm for 40 minutes. Ni-NTA resin was equilibrated (Hepes 20 mM pH 7.8, NaCl 200 mM) in preparation for the affinity chromatography, and two hours binding at 4° C were done. The column was washed with three different buffers respectively: 20 mL washing buffer 1 (Hepes 20 mM pH 7.8, NaCl 200 mM), 20 mL washing buffer 2 (Hepes 20 mM pH 7.8, NaCl 500 mM), 20 mL washing buffer 3 (Hepes 20 mM pH 7.8, NaCl 350 mM, imidazole 75 mM), and the proteins fractions were recovered in 10 mL elution buffer (Hepes 20 mM pH 7.8, NaCl 350 mM, imidazole 500 mM). Finally, TEV was dialysed in Hepes 20 mM pH 7.8, NaCl 150 mM, Dithiothreitol (DTT) 1mM, glycerol 30%.

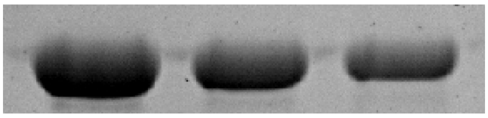


Figure 3.2: Tobacco etch virus protease purification. Detection of recombinant Tobacco etch virus protease in SDS-PAGE after the purification performed with affinity chromatography with nickel resin. Here are represented three elutions of the protein that were collected separately and then pulled together.

3.1.3 Protease cleavage

The N-term six histidine tags of aIF5A were cut using TEV protease 1:50 ratio. The cleavage reaction took place overnight at room temperature in the following buffer: Tris-HCl pH 8 500 mM, ethylenediaminetetraacetic acid (EDTA) 5 mM, DTT 10 mM. The digestion was accurately investigated in Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE). The affinity chromatography was realized using Ni-NTA resin, which was equilibrated with KCl 150 mM and Tris-HCl pH 8 50 mM, and washed using the same buffer at increasing concentration of KCl, respectively: 150 mM, 300 mM and 500 mM. aIF5A was found in in flowthrough and then dialysed in Tris HCl pH 8.5 50 mM, NaCl 150 mM and glycerol 5%. Fig 3.3 reports the cleavage of aIF5A before and after the reaction with TEV.

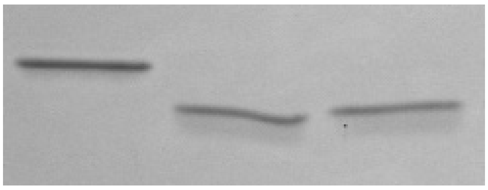


Figure 3.3: The bands shown in figure represent aIF5A in SDS-PAGE before and after the cut of its histidine tag. The first bar shows aIF5A with its histidine tail (17 kDa), and the other two elutions represent aIF5A after the cleavage (15 kDa).

3.1.4 Recombinant C-term his tag aDHS purification

The purification of aDHS in pqE70 plasmid inserted in *E. coli* BL21 was done through a standard protocol already optimized (Fig. 3.4 shows the SDS-PAGE of the protein.). The culture medium was the same used for aIF5A, but in this case just ampicillin was present as antibiotic. The induction with IPTG 1 mM (4 hours) was done once the optical cells density reached values between 0.4 and 0.8, and then cells were collected by centrifugation. Chemical lysis was induced with lysis buffer (Tris HCl pH 6-6.8 50 mM, KCl 150 mM, imidazole 15 mM, PMSF 1 mM, β-Mercaptoethanol 10 mM, lysozyme 25 μg/mL, triton X 0.1%) and followed by sonication and centrifugation at the same condition of the previous purification. The purification with Ni-NTA Agarose resin, previously equilibrated (Tris-HCl 6-6.8 50 mM, NaCl 150 mM, imidazole 15 mM) was carried out using 50 mL washing buffer (Tris-HCl pH 6-6.8 50 mM, NaCl 1 M, imidazole 40 mM) and the protein recovered in 5 mL elution buffer (Tris-HCl 6 50 mM, NaCl 150 mM, imidazole 250 mM). Finally, the protein was dialysed in dialysis buffer (Trish HCl pH 8.5 50 mM, NaCl 500 mM, imidazole 80 mM and glycerol 5%). A small amount of imidazole was left in the sample because recent experiments have underlined its capacity to stabilize this protein.

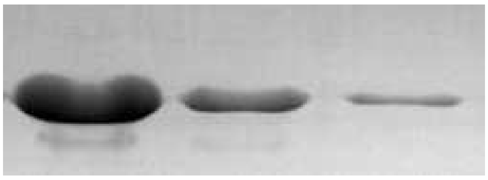


FIGURE 3.4: Detection of recombinant aDHS protein in SDS-PAGE after the purification performed with affinity chromatography with nickel resin. The three bars represent the elutions of the protein recovered separately and then pulled together.

3.1.5 Recombinant aIF5A and aDHS complex

The complex between the two proteins was obtained in Tris-HCl pH 8.5 50 mM, NaCl 150 mM, imidazole 50 mM and glycerol 5%. aIF5A and aDHS were mixed with a molar ratio of 6:1, respectively, and then the complex was formed. In Fig. 3.5 you can see the SDS-PAGE of the complex.

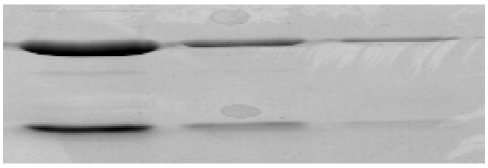


FIGURE 3.5: SDS-PAGE of the complex between aIF5A without his-tag and aDHS. The highest three bands indicate the presence of aDHS as a monomer in the elutions, whereas the lowest three lines represent aIF5A (15 kDa).

3.1.6 SEC-SAXS and Elettra setup optimization

To get a complete set of information about the sample under analysis, it is necessary to rely on complementary techniques. In modern synchrotrons MALS, DLS, UV-VIS, and Refractive Index detector (dRI) can be utilized, together with SAXS, in order to extract information about molecular dimension, molar mass and hydrodynamic radius. The SEC-SAXS setup built at the Austrian SAXS beamline at the Elettra Synchrotron of Trieste [3] is composed by a pump, an autosampler, a chromatographic column as well as by UV-VIS, MALS and dRI detectors (HPLC system: Agilent, Santa Clara, CA 95051, USA; MALS/qUELS, dRI: Wyatt Technology, Santa Barbara, CA 93117, USA) and all these instruments are connected in series with the SAXS cell. The SEC apparatus with all its components is shown in Fig. 3.6.

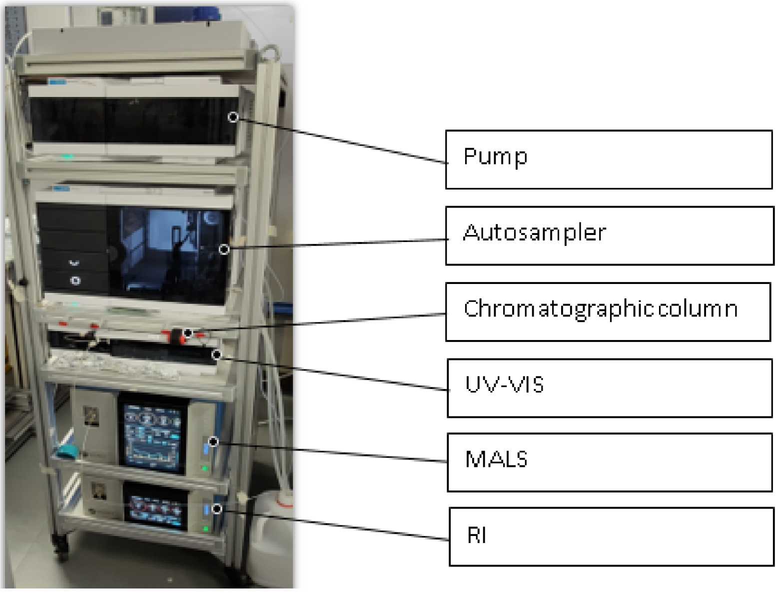


Figure 3.6: Size Exclusion Chromatography apparatus recently installed at Elettra.

In Molecular Biology, the importance of using SEC prior to SAXS is mostly related to the type and the aggregational behaviour of the sample that needs to be analysed. If a protein is mainly present in solution as an aggregate, the use of Size Exclusion Chromatography becomes fundamental to separate the different oligomers according to their size, allowing to measure the SAXS curve of each species.

In synchrotrons SAXS beamlines, different ways of building up the SEC-SAXS setup have been experimented. As an example, a splitter valve can be used to separate the flux into two portions after the chromatographic column and before the UV-VIS detector [30]. In this setup, part of the sample undergoes UV-VIS, MALS, and dRI analysis, and the rest, carefully synchronised, is transported to the SAXS observation cell for X-rays scattering analysis. At the Elettra SAXS beamline, the same setup was tested, but no satisfactory results were obtained. The main problem was the presence of air bubbles in MALS detector due to the splitter itself that was found to suck in air. The synchronization of the two different fluxes was also not easy, in particular considering the appropriate flowrate to use. As previously mentioned, in order to have a high separation of molecules a low flowrate is essential, but if the flowrate is too low, the MALS peaks broaden reducing the resolution. In addition, the flowrate in the SAXS cell should be low enough to guarantee a sufficient exposure time. During the experiments performed at the Austrian SAXS beamline, a minimum of 10 s per exposure was necessary to obtain a good SAXS signal from diluted protein samples: adjusting the ratio between the SAXS and MALS lines did not give satisfactory results, so the splitter option was abandoned and different configurations were tested, in order to optimize the system. I took an active part in the optimization of the setup spending four months at the Elettra synchrotron. To have the maximum concentration of the sample to optimize SAXS signal, the SAXS cell was connected directly to the chromatographic column prior to the UV-VIS detector, and in order to understand the system dilution effect, connection tubes of different inner diameters were tested. For all the investigations, Bovine Serum Albumin (BSA) at a concentration of 5 mg/mL was chosen as model protein, and Superdex 75 10/300 (Cytiva) was the chromatographic column used for the analysis. Four different configurations were set up and the degree of dilution was calculated using UV-VIS. The first configuration was taken as a reference to have a clear idea about the protein concentration in absence of the SAXS cell. Afterward, several tube lengths were prepared with different diameters and connected to the capillary. In Fig. 3.7, two tubes are represented: the green wire has an internal diameter of 0.8 mm and the one of the orange wire is 0.125 mm. In Table 3.1, the detailed features of the tubes are described.

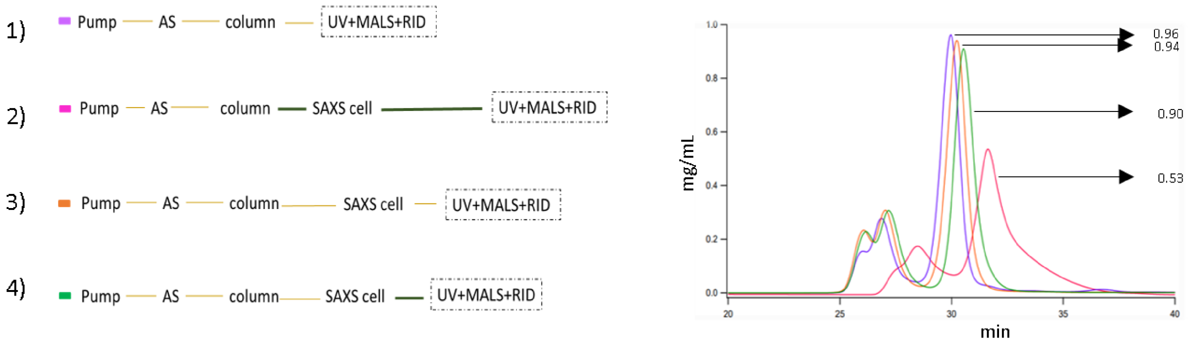


Figure 3.7: On the right the chromatograms of BSA are reported. The colours of the curves indicate the different configurations tested, whose schematic description is reported on the left and highlighted by the respective coloured rectangle. All the configurations are detailed explained in the table below 3.1, where the dimensions (length and diameter) of the connected tubes used for each analysis is reported.

Configuration	**Inlet tube length (cm)**	**Outlet tube length (cm)**	**Internal diameter (ID) (mm)**
1 no SAXS cell	/	/	0.125
2	5	11	0.8
3	39	21	0.125
4	39	15	0.125 and 0.8

Table 3.1: Elettra SEC-SAXS configurations.

Looking at Fig. 3.7, it is clear how the setup configuration influences the dilution of the sample starting with a basic dilution factor that is not referred to the SAXS cell but to the system itself (the purple curve, configuration 1). Here no cell is installed and therefore the spectrum, obtained by UV-VIS, shows the highest peak in accordance with the highest protein concentration. The insertion of the SAXS cell and of the connection tubes gradually increases the dilution factor. The biggest ID tube (0.8 mm, configuration 2) leads to the maximum protein dilution as it is highlighted by the pink curve where the protein concentration reaches a dilution factor of ten (configuration 2). Considering the plots represented in Fig. 3.7, the best choice should be the third configuration, where two connection tubes of ID 0.125 were used. However, after one hour experiment the capillary broke because of the back pressure exerted by the thin outlet tube and by the following instruments against the SAXS cell walls, and for this reason this cannot be considered the optimum setup. Other tests were done using a tube with ID of 0.8 mm as an outlet tube to connect the SAXS cell to MALS, but breaks and leaks of the capillary were present anyway (configuration 4). After careful examinations, the best configuration appeared to be the one with the capillary at the end of the entire setup accurately connected with the orange tube, which in the end brings to a dilution factor of five. For this reason, proteins need to be highly concentrated (at least 10 mg/mL) to give a clear signal in the SAXS instrument.

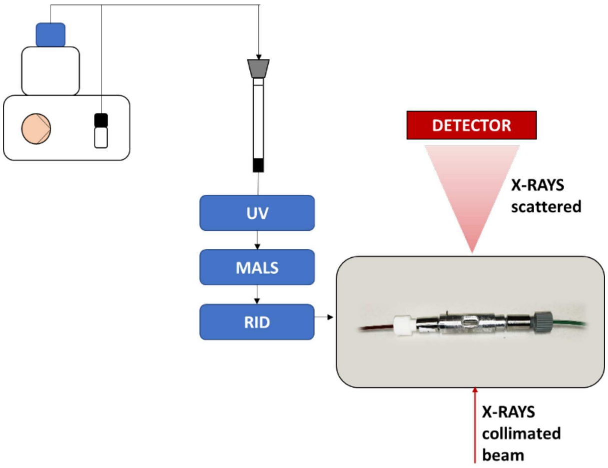


Figure 3.8: Optimum setup configuration.

Inlet tube (mm)	Outlet tube (mm)	Capillary material	Capillary wall thickness (mm)	Capillary internal thickness (mm)
0.125	0.88	quartz	0.01	1

Table 3.2: Features of the optimum configuration found.

3.1.7 SEC-SAXS experiment and sample preparation

Using the recently installed and optimized SEC-SAXS apparatus at the Austrian SAXS beamline of Elettra (Fig. 3.9), the oligomeric distribution of aIF5A has been investigated, as well as the distribution of aDHS and of the aDHS-aIF5A complex. In order to obtain detailed results, Superdex 200 10/300 GL was the chromatographic column used, which allows molecules from 10 to 600 kDa to be accurately separated and therefore analysed. Protein solutions were filtered with a 0.2 μm pore size filter and then 100 μL of each protein solution was loaded into the column. According to the different protein concentration in the three samples, the total mass of the inserted proteins corresponded to 920 μg of aIF5A, 180 μg of aDHS and 320 μg of the aDHS-aIF5A complex without histidine. A constant flowrate of 0.3 mL/min was set for all the investigations at 20° C, and the SAXS setup was fixed with the optimum q range around $0.01 - 0.5$ Å^{-1}.

Protein	**Concentration (mg/mL)**	M_w **(kDa)**	**Buffer**
aIF5A	9.2	14.5 monomer	Tris HCl pH 8.5 50 mM, NaCl 150 mM, glycerol 5%
aDHS	1.8	144 tetramer	Tris HCl pH 8.5 50 mM, NaCl 500 mM, imidazole 80 mM, glycerol 5%
Complex	3.2	Around 200	Tris HCl pH 8.5 50 mM, NaCl 150 mM, imidazole 50 mM, glycerol 5%

Table 3.3: List of proteins analysed and their features.

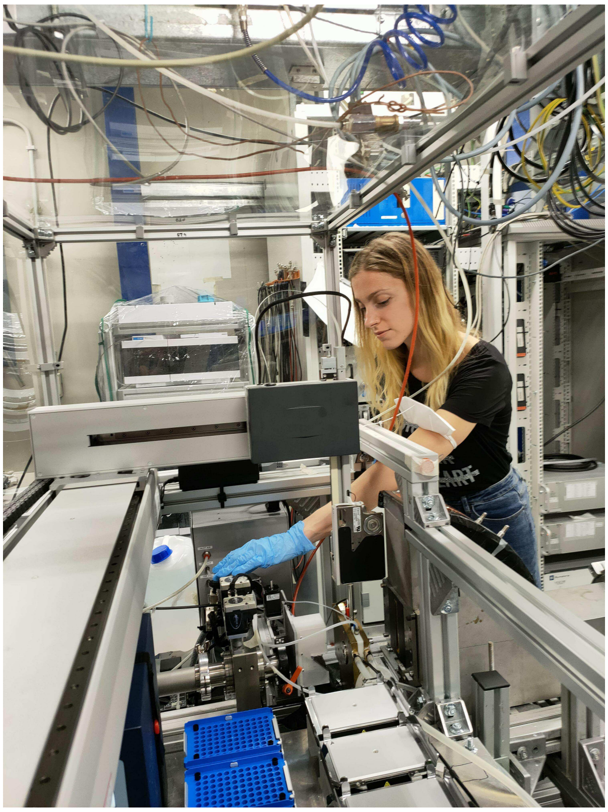


FIGURE 3.9: AustroSAXS beamline at the Elettra synchrotron in Trieste.

3.1.7.1 aIF5A chromatograms

The software used for the characterization of particles undergoing size exclusion chromatography was developed by Wyatt Technology Corporation and it is called ASTRA. By means of this software, the chromatograms derived by UV-VIS, MALS, and dRI were analysed and information about the molecular weight of the particles in the solution was obtained. The red peak shown in Fig. 3.10 represents aIF5A in its monomeric state, as it is confirmed by ASTRA, underlying that the majority of the protein is present in solution as a monomer.

Then, by using the software IGOR Pro, it was possible to average the SAXS frames associated to the chromatography peak attributed to the monomer and subtract the SAXS

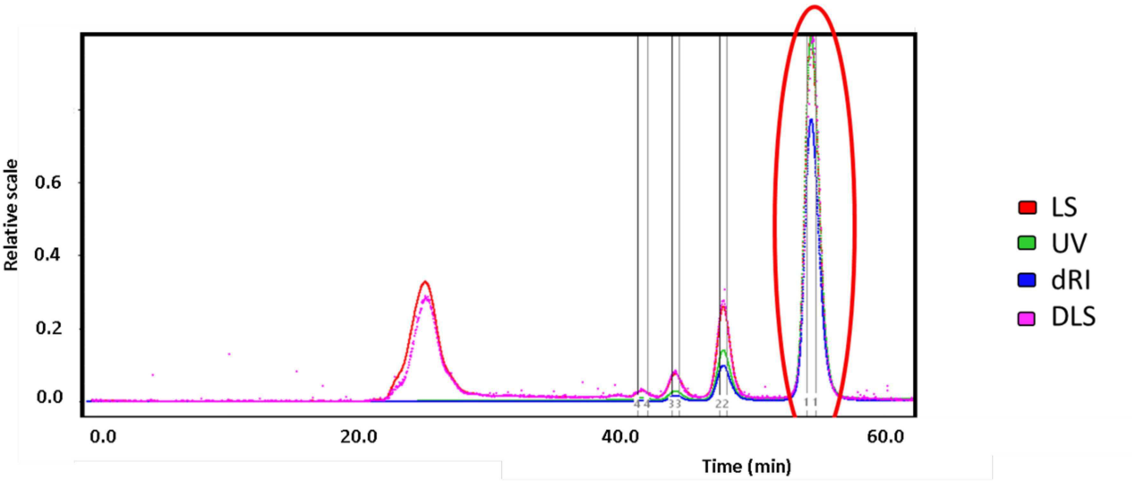


Figure 3.10: The graph shows the chromatograms of aIF5A derived from all the detectors and elaborated by ASTRA software.

frames related buffer in order to obtain the SAXS profile of the monomer in solution, shown in Fig. 3.11. The broad peak shown in the Kratky plot in Fig. 3.12 underlies that the compactness of the protein in solution is mostly maintained, with no loss of its native conformation.

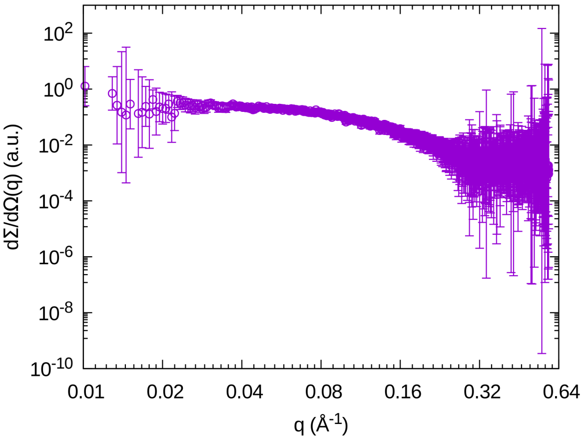


Figure 3.11: SAXS curve of aIF5A obtained after the SEC analysis. This curve derives from the analysis of the monomer highlighted in Fig. 3.10 and was obtained by averaging the SAXS frames in correspondence to the chromatographic peak and by subtracting the average SAXS frames attributed to the buffer.

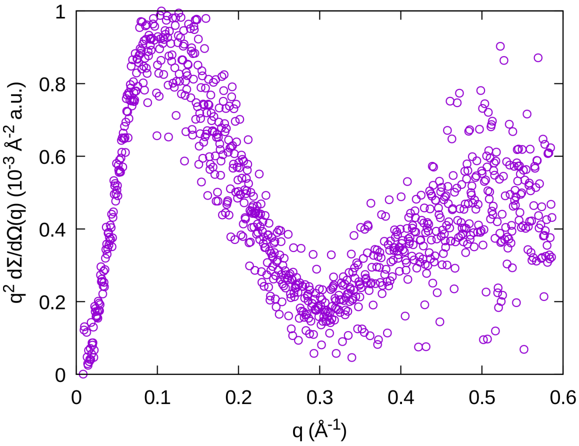


FIGURE 3.12: The Kratky plot of aIF5A showing a broad peak typical of compact particles

3.1.7.2 Complex aDHS-aIF5A chromatograms

The same procedure was applied to determine the SEC-SAXS curve of the complex between aDHS and aIF5A. The chromatograms, the SAXS curve and the corresponding Kratky plot are shown in Figs. 3.13, 3.14 and 3.15, respectivley.

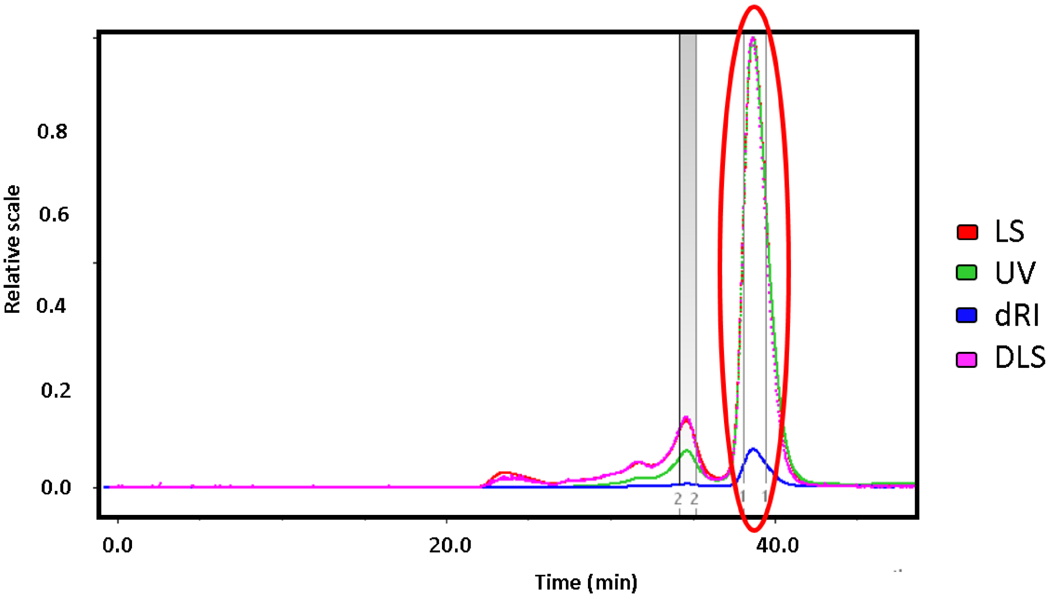


FIGURE 3.13: The graph shows the chromatograms of the complex derived from all the instruments and recorded by ASTRA software.

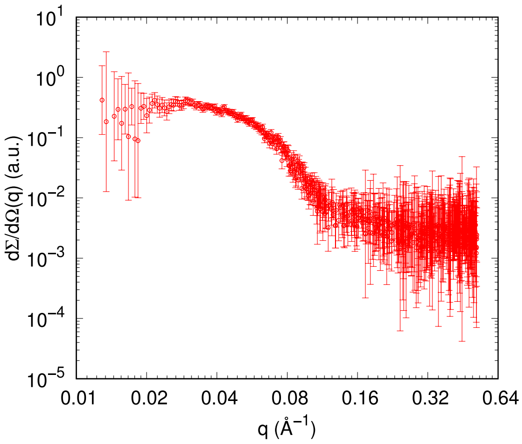


FIGURE 3.14: Complex of aIF5A and aDHS.

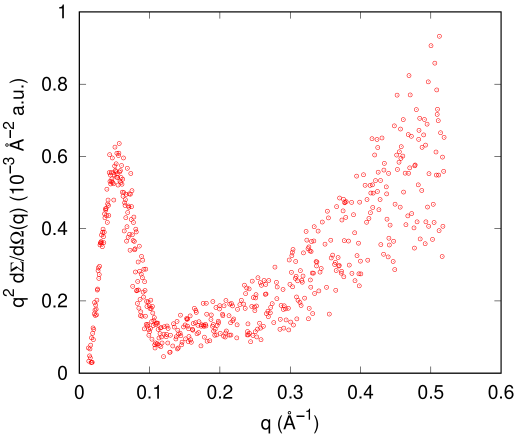


FIGURE 3.15: The Kratky plot of the complex of aIF5A and aDHS.

3.1.8 SEC-SAXS data analysis

The SAXS curve of aIF5A, in its monomeric form, resulting from SEC-SAXS experiment, has been analyzed by using the GENFIT software, developed in the Molecular Biophysics Laboratory of the Department of Life and Environmental Sciences (DISVA) of the Polytechnic University of Marche (Ancona, Italy) [81]. The analysis has been planned considering that, being the protein very diluted, the SAXS curve mostly depends on the monomer form factor, according to Eq. 3.9. However, since the protein has not yet been crystallized, the atomic structure of the aIF5A protein of *S. solfataricus*, from which its form factor can be calculated, is not available. We have then analyzed the SEC-SAXS curve of aIF5A by taking into consideration a pool of $N = 19$ homologous atomic structures of IF5A adopted in a recent study [69]. Six of these structures belong to Archaea and the others derive from Eukarya. The corresponding and available 19 PDB files have been processed with the SASMOL approach [58] (a method included in GENFIT), to get their SAXS form factor $P_j(q)$, which are shown in Fig. 3.16.

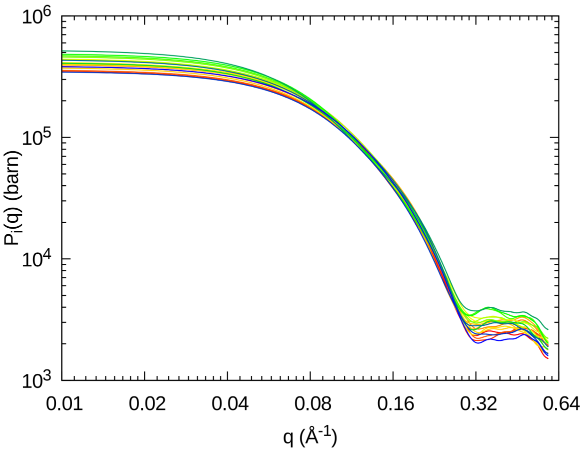


FIGURE 3.16: Log-log plot of the form factors of the 19 structure employed to fit the SEC-SAXS curve of *S. solfataricus* aIF5A.

To quantitatively estimate the likelihood of each of these 19 form factors with the form factor of *S. solfataricus* aIF5A, the SAXS experimental curve has been analysed as a weighted average of $P_j(q)$, according to the following equation

$$\frac{d\Sigma}{d\Omega}(q) = \kappa \sum_{j=1}^{N} w_j P_j(q) + B, \tag{3.1}$$

where the fitting parameter w_j represents the relative weight of the j-structure, with the normalization condition

$$\sum_{j=1}^{N} w_j = 1. \tag{3.2}$$

Notice that in Eq. 3.1 the other two less important fitting parameters, κ and B, represent a scaling factor and a flat background, respectively, both necessary since, from one hand, the SEC-SAXS curve was not given in absolute scale and, from the other hand, the buffer subtraction may have been affected by an unspecified error. The best fit is achieved by GENFIT via the minimization of the following functional,

$$\mathcal{H} = \chi^2 - \gamma \hat{S} \tag{3.3}$$

where χ^2 is the standard reduced chi-squared of the experimental curve $\frac{d\Sigma}{d\Omega}_{\text{exp}}(q_k)$ recorded over N_q values of q,

$$\chi^2 = \frac{1}{N_q} \sum_{k=1}^{N_q} \left(\frac{\frac{d\Sigma}{d\Omega}_{\text{exp}}(q_k) - \frac{d\Sigma}{d\Omega}(q_k)}{\sigma_{\text{exp}}(q_k)} \right)^2 . \tag{3.4}$$

Here, $\sigma_{\text{exp}}(q_k)$ is the experimental uncertainty (estimated by the standard deviation over all the frames considered for the calculation of $\frac{d\Sigma}{d\Omega}_{\text{exp}}(q_k)$) and $\frac{d\Sigma}{d\Omega}(q_k)$ is the best fitted curve calculated with Eq. 3.1. The other term of the functional $\mathcal{H}$ appearing in Eq. 3.3 is the *statistical entropy* associated to the set of w_j (that we collectively call *population weights*),

$$\hat{S} = -\sum_{j=1}^{N} w_j \log w_j . \tag{3.5}$$

On the basis of Eq. 3.3, the minimization of $\mathcal{H}$ involves not only the minimization of χ^2 but also the maximization of $\hat{S}$, according the more general *maximum entropy principle*, and this will provide the so-called *least biased* population weights distribution. The constant γ in Eq. 3.3 is carefully chosen in order to ensure that, at the end of the minimization procedure, the factor $-\gamma\hat{S}$ accounts for $\approx 10\%$ of $\mathcal{H}$, hence by guaranteeing that the best fit of the data (i.e. the χ^2 term) is the dominant factor. In Fig. 3.17 we plot the fitted curve obtained with GENFIT: we can appreciate the goodness of the fit.

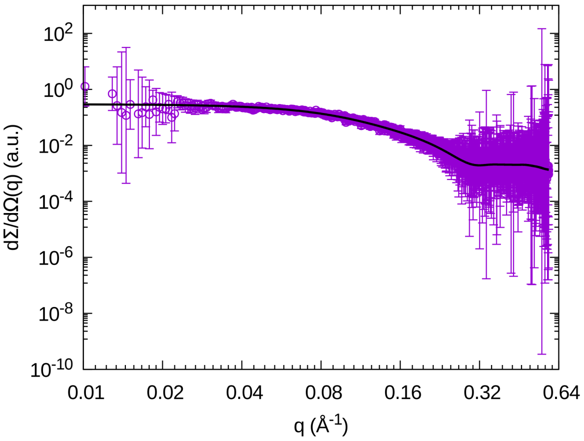


Figure 3.17: Best fit of the curve shown in Fig. 3.11.

The obtained population weights are shown, in the form of an histogram with decreasing value of w_j, in Fig. 3.18.

By looking at the Fig. 3.18 it is clear that the main probability of the translation factor IF5A, derived from the archaea organism *S. solfataricus*, is to be similar to the archaea structures in respect to the eukaryotic ones, as it was expected. Hence, by using the approch described above and synthesised in the equation 3.1 we confirmed that the recombinant protein used in this work has a structure deeply similar to other archaea organisms and in particular to another thermophilic organism called *Pyrococcus horikoshii.*

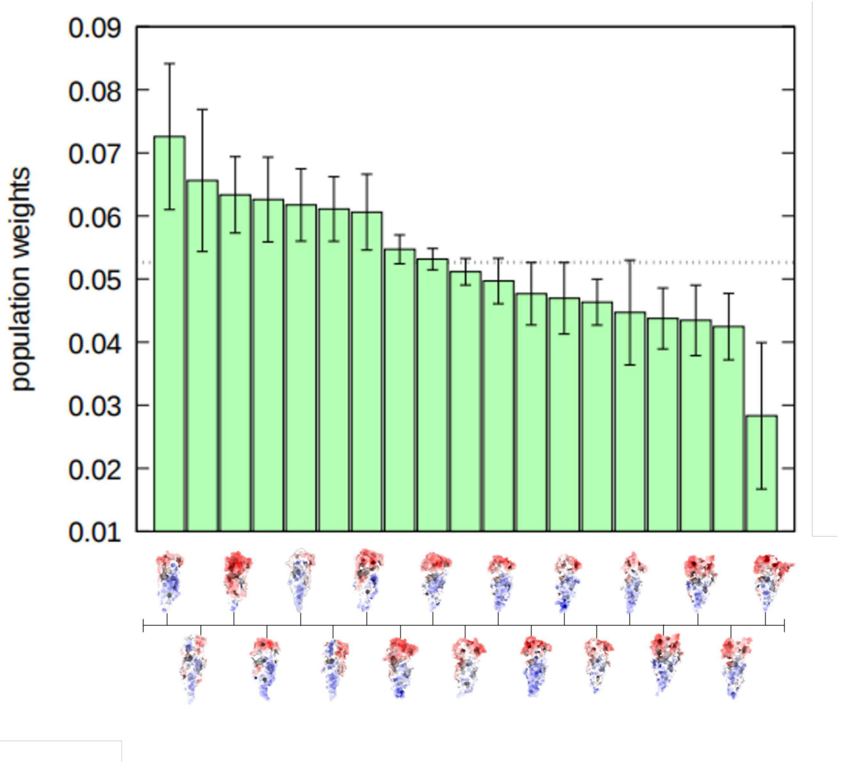


Figure 3.18: Population weight distribution of the 19 structures of IF5A deriving from both Archaea and Eukarya domains determined by the best fit of the SEC-SAXS curve of the recombinant aIF5A deriving from *S. solfataricus* and expressed in *E. coli*.

The same fitting strategy was adopted for the analysis of the SEC-SAXS curve of the complex formed by aDHS and aIF5A, but in this case the aim was to decipher the stoichiometry of the complex in solution. To achieve this goal we have employed the atomic structure of the complex that has been recently found by using X-ray crystallography (data unpublished, given by courtesy of Mattia D'Agostino, Fig 3.19).

In this structure, there are 4 aIF5A molecules attached to each of the 4 sub-unit of the DHS tetramer. It is therefore necessary to verify, on the basis of the SEC-SAXS experiments, whether the crystallographic structure of this complex is present in solution, or whether all the other possible structures formed by the combination of a different number of aIF5A molecules with DHS may also be present. To this end, we selectively

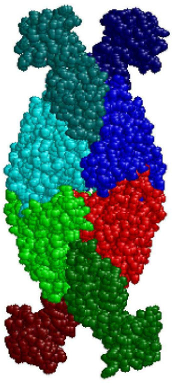


Figure 3.19: View of the X-ray crystal structure of the complex aIF5A and aDHS obtained by Dr. Mattia D'Agostino. Green, red, blue and cyan domains (A-D chains in the PDB file) are the 4 sub-unit of the compact aDSH tetramer, each of which bounds one aIF5A molecule, represented in dark-green, dark-red, dark-blue and dark-cyan (chains E-H in the PDB file).

removed one or more aIF5A molecules from the PDB file of the 4:1 complex (containing a total of 8 chains, the first 4 (A to D) related to the 4 sub-units of DHS and the other 4 (E to H) corresponding to the 4 aIF5A molecules bound, in the same order, to the A-D chains of DHS), and thus generating 6 possible structures. They correspond to the A-D,E sequence (1:1 complex), the A-E,F, A-E,G and A-E,H sequences (three distinct 2:1 complexes), the A-G sequence (3:1 complex) and the entire A-H sequence (original 4:1 complex). A representation of the 6 complexes is shown in Fig. 3.20.

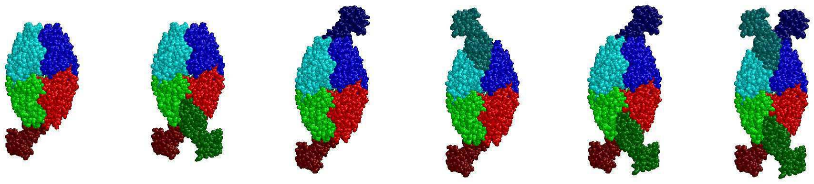


Figure 3.20: Comparison between the crystalized complex composed of four monomers of aIF5A, and other structures that have been hypothesised to be present in solution.

The form factors of these structures have been calculated with the SASMOL approach. They are plotted as a function of q in Fig. 3.21.

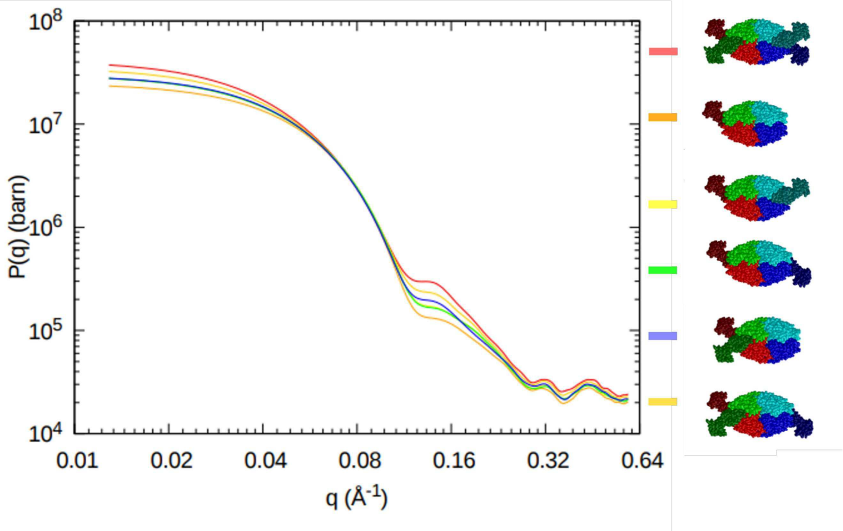


Figure 3.21: Form factor of all the hypothetical representations of the complex in solution.

The experimental SEC-SAXS curve of the complex aIF5A-aDHS (Fig. 3.14) has been then analysed with the same approach used of the analysis of aIF5A (Eqs. 3.1-3.5), by using the $N = 6$ form factors shown in Fig. 3.21. The best fit curve is shown in Fig. 3.22 and the obtained population weight distribution is reported in Fig. 3.23.

By looking at the Fig. 3.23 we can appreciate that the most likely configuration is not the one of the crystallized structure as it could be expected. On the contrary, the configuration of the complex surrounded by four monomers of IF5A is the least likely to occur, whereas the two protein species seem to be mainly present in solution as a sum of the ezyme with two monomers of aIF5A in three different positions.

3.1.9 The aIF5A conclusions

We have shown that, with a detailed method that uses SEC prior to SAXS analysis, it is possible to extract detailed structural information regarding the preferential state of

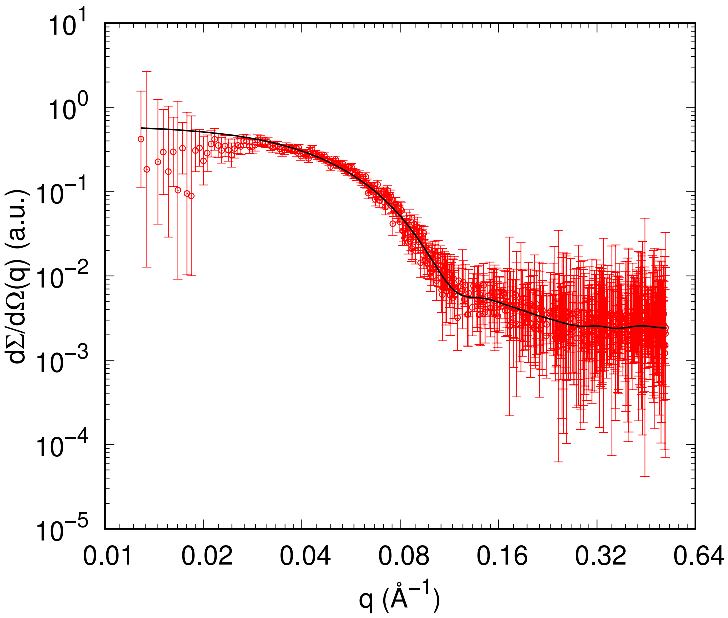


Figure 3.22: Fitting of the form factor with the experimental curve of the complex derived from SEC-SAXS.

proteins in solution. The protein production and purification, which took my first year of PhD, was the first necessary step to get all the protein species for the following studies. A key role was played by the optimization of the SEC apparatus at the SAXS beamline of the Elettra synchrotron, which has never been used before our experiment. The different tests that we did for several months to find the right configuration of the system were of pivotal importance in order to develop the best SEC-SAXS design with less breaks and leaks of the capillary. Moreover, the possibility of studying a single monomeric state of the aIF5A protein was crucial for our experiment, in particular considering the tendency of the archaeal translation factor to aggregate, and the difficulty to obtain a Gaussian-like trend (Eq. 2.13) in the low q part, as it was found on the SAXS curves in previous experiments [69].

The simulation of the scattering profile of the proteins derived from different organisms,

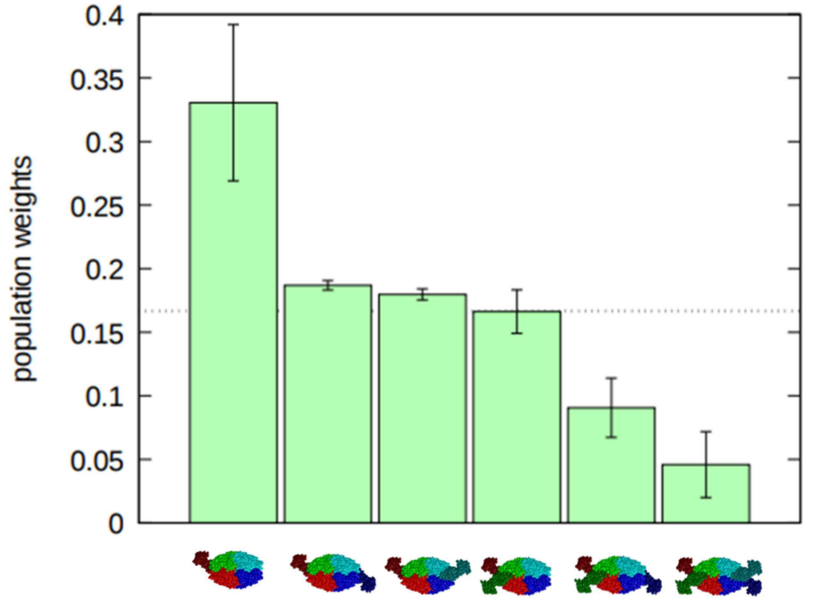


FIGURE 3.23: Population weight distribution of the 6 possible complexes formed by aIF5A and aDHS determined by the best fit of the SEC-SAXS curve shown in Fig. 3.22.

achieved using the software SASMOL and exploited by the GENFIT software, allowed the deduction of the structure of aIF5A protein in solution. The result obtained was a good confirmation that the protein from *S. solfataricus* is more similar to the one derived from another termophilic organism: *Pyrococcus horikoshii*.

Another important point that was clarified in this thesis was the preferential configuration in solution of the complex between aIF5A and aDHS. The crystallized structure of the complex 3.19 is composed of four monomers of aIF5A linked to each of the four subunits of aDHS, and this is probabily due to a matter of stability and geometry needed during the crystallization procedure. However, since there can be some differences between proteins in solution and proteins crystallized we decided to investigate the conformation of the complex in solution applying the SEC-SAXS analysis. By means of the same strategy adopted for the analysis the SEC-SAXS curve of aIF5A based on both SASMOL and

GENFIT software, we have quantitatively estimated the probability to find, in solution, the complex in different configurations states. The configurations took into account are reported in Fig. 3.20 and among them, the preferential state of the complex in solution appeared to be the one composed of two monomers of aIF5A instead of four, even if also the probability to find the complex with just one monomer of the translation factor is high, but not as high as the one derived by the sum of two monomers of aIF5A at three different positions.

3.2 SAXS studies of protein stabilization by modified sugars

3.2.1 SAXS experiment and samples preparation

Myoglobin (MB) from equine heart 2 and 10 g/L was dissolved in 10 mM phosphate buffer at pH 5, and insulin (IN) 2 g/L from bovine pancreas was prepared in the same buffer at pH 3. The Portuguese chemical synthesis company ExtremoChem developed five modified sugars (EC101, EC202, EC212, EC311, EC312) with moiety composed of glucose, mannose, or galactose and charged or neutral substituents at the anomeric position. Moreover, two sugars (EC101 and EC202) when dissolved in water behave like ionic species and slightly increase the pH. The experimental plan is descibed in tables below 3.4 3.5. The q scattering vector was fixed between 0.01 and 0.35 $\mathring{A}^{-1}$.

MB concentration (g/L)	Buffer	Modified sugar	Modified sugar concentration (M)	Temperature ° C
2 10	10 mM phosphate pH 5	EC101 EC202 EC212 EC311 EC312	0.05 0.1 0.25	25 35 60

Table 3.4: Myoglobin experimental plan.

The following paragraphs have been obtained from the article that we published in January 2022 in *Life* journal: *SAXS reveals the stabilization effects of modified sugars on model protein* [63]

IN concentration (g/L)	Buffer	Modified sugar	Modified sugar concentration (M)	Temperature ° C
2	10 mM phosphate pH 3	EC101 EC312	0.05 0.1 0.25	25 30 35 40 45 55 60

Table 3.5: Insulin experimental plan.

3.2.2 SAXS Experiments

SAXS experiments were performed at the Austrian SAXS beamline of Elettra synchrotron (Trieste, Italy) [3]. Measurements were carried out at 25°, 35° and 60° C for myoglobin, whereas the same analysis was performed at 25°, 30°, 35°, 40°, 45°, 55° and 60° C for insulin. During the whole experiment the new μ-Drop sample changer recently developed in the Austrian beamline was used [32]. The modulus q of the scattering vector, related to the scattering angle 2θ and to the x-ray wavelength $\lambda = 1.54$ Å by the relationship $q = (4\pi/\lambda)\sin\theta$, was fixed between 0.01 to 0.35 Å^{-1}. For each sample, twelve bidimensional and isotropic SAXS patterns were collected by a Pilatus3 1M detector and subsequently treated with FIT2D [33] to apply the beamstop and detector mask and to perform the radial average. Finally, by using the SAXS data reduction system (SAXS dog) for the subtraction of the buffers isotropic SAXS signal from the one of the samples, the normalization to the intensity of the primary beam and the correction for the samples' transmissions, the experimental macroscopic scattering cross section $\frac{d\Sigma}{d\Omega}(q)$ of each sample was obtained.

3.2.3 SAXS Data Analysis

The analysis of SAXS data has been performed by assuming that proteins in solution can be present in N_s different *states* (e.g. folded oligomers or unfolded chains), without any preferential orientation, and that long-range isotropic protein-protein interactions may occur. In these circumstances, the macroscopic differential scattering cross section (the precious information provided by SAXS experiments), shown in Eq. 2.8, can be expressed as

$$\frac{d\Sigma}{d\Omega}(q) = n_1 \, P(q) \, S_M(q). \tag{3.6}$$

where n_1 is the nominal number density of the protein monomers, simply related to the w/v protein concentration, c, through Avogadro's number, N_A, and the monomer molecular weight, M_1, by $n_1 = cN_A/M_1$. The second term, $P(q)$, already defined as the *average form factor* in Eq. 2.10, can be written as

$$P(q) \quad = \quad \sum_{j=1}^{N_s} \frac{x_j}{\alpha_j} P_j(q) \tag{3.7}$$

where $P_j(q)$ is the form factor (the orientational average of the squared excess X-ray scattering amplitude) of the j-protein state, α_j is the corresponding aggregation number, whereas x_j is the molar fraction of nominal protein monomers that are forming the j-state, with the condition

$$\sum_{j=1}^{N_s} x_j = 1 \tag{3.8}$$

The third term, $S_M(q)$, known as the *measured structure factor* (see section 2.1.3.1), depends on the average protein-protein structure factor, $S(q)$, according to $S_M(q) = 1 + \beta(q)[S(q) - 1]$, where $\beta(q)$ is the coupling function, $\beta(q) = |P^{(1)}(q)|^2/P(q)$, $P^{(1)}(q)$ being the weighted average of the orientational average of the excess X-ray scattering amplitude $P_j^{(1)}(q)$ of the j-state

$$P^{(1)}(q) = \sum_{j=1}^{N_s} \frac{x_j}{\alpha_j} P_j^{(1)}(q). \tag{3.9}$$

The calculation of both $P_j(q)$ and $P_j^{(1)}(q)$ has been carried out on the basis of protein data bank (PDB, [13]) atomic structure associated to the j-state by using the SASMOL approach [58]. This method is based on the description of the solvent molecules in contact with the protein as dummy Gaussian spheres and determines the number and the geometrical coordinates of such spheres by burying the protein in a tetrahedral close-packed (TCP) lattice of dummy spheres. Consequently, the number and the positions of the water molecules can be obtained in the first N_{sh} hydration shells of the j-protein state and a scattering length density (SLD) that can differ from the one of the bulk solvent is assigned to each of them. Typically, the thickness of each water shell is considered to be equal to 2.8 Å. Notice that in this work we have considered $N_{sh} = 2$. This feature is particularly useful in the presence of a binary solvent, such as a solution of water and modified-sugar, where preferential solvation effects can lead to a modification of the composition of the binary solvent in contact with the protein surface with respect to the composition of the bulk binary solvent.

3.2.3.1 Multimeric equilibrium processes in binary solvents

In equilibrium conditions, the distribution of proteins in the N_s states and the composition of the first protein hydration shell as a function of protein concentration, solvent composition and temperature can be determined by considering the interplay of different elementary processes. *First*, we consider the process of transformation of a protein (hereafter indicated by the symbol P) dissolved in water at a certain pH and at a certain ionic strength (I) from the state 1 (which is assumed to be a monomeric state, typically a native state) to the state j,

$$\mathrm{P}_1(\mathrm{W}_{\mathrm{s}1})_{m_1} \rightleftharpoons \alpha_j^{-1}\mathrm{P}_j(\mathrm{W}_{\mathrm{s}j})_{m_j} + (m_1 - \alpha_j^{-1}m_j)\mathrm{W}_\mathrm{b} \tag{3.10}$$

where m_j is the number of water sites in the first hydration shell of the j-state (a value that can be determined by SASMOL), $\mathrm{W}_{\mathrm{s}j}$ represents a water molecule attached to the surface of the protein in the j-state and W_b represents a water molecule in the bulk (see Fig. 3.24). By assuming an ideal thermodynamic behavior of the system, the equilibrium constant $K_{\mathrm{W}1j}$ as well as the standard Gibbs free energy change $\Delta G_{\mathrm{W}1j}$ associated to this process is

$$K_{\mathrm{W}1j} = \frac{C_{\mathrm{P}_j(\mathrm{W}_{\mathrm{s}j})_{m_j}}^{\alpha_j^{-1}} X_{\mathrm{W}_\mathrm{b}}^{m_1-\alpha_j^{-1}m_j}}{C_{\mathrm{P}_1(\mathrm{W}_{\mathrm{s}1})_{m_1}}} = e^{-\Delta G_{\mathrm{W}1j}/(\mathrm{R}T)} \tag{3.11}$$

where the symbol C is the molar concentration (used for the solutes, being $C = 1$ M their standard state) and the symbol X stands for the molar fraction (used of the solvent

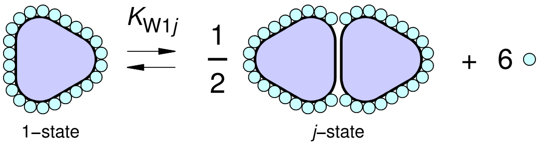


FIGURE 3.24: Sketch of an equilibrium process of the protein in water from the monomeric (1-)state to the dimeric (j=2) -state (scheme 3.11, with $\alpha_j = 2$ and $m_1 - \alpha_j^{-1} m_j = 6$).

(water), being $X = 1$ its standard state). The *second* process refers to proteins dissolved in a binary solvent constituted by water and a cosolvent (such as a modified-sugar) and describes the exchange of a cosolvent molecule attached to the first hydration shell of the protein in the j-state (indicated by the symbol $\mathrm{G}_{\mathrm{s}j}$) with a bulk water molecules (see Fig. 3.25 for a clarifying example),

$$\mathrm{G}_{\mathrm{s}j} + \mathrm{W}_\mathrm{b} \rightleftharpoons \mathrm{G}_\mathrm{b} + \mathrm{W}_{\mathrm{s}j} \tag{3.12}$$

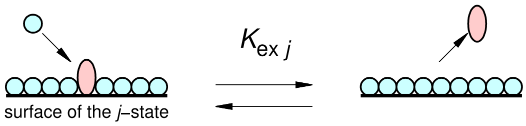


FIGURE 3.25: Sketch of the water-cosolvent (blue spheres and red ellipsoids, respectively) exchange equilibrium process over the surface of the j-protein state.

that leads to the formation of a cosolvent molecule in the bulk (G_b) and a water molecule in the first shell ($\mathrm{W}_{\mathrm{s}j}$). According to the well-established Schellmann model [72, 73, 74, 75], this *exchange equilibrium* has been found to be simply described by the thermodynamic constant $K_{\mathrm{ex}j}$ and the related standard Gibbs free energy change $\Delta G_{\mathrm{ex}j}$,

$$K_{\mathrm{ex}j} = \frac{\phi_j X_{\mathrm{G_b}}}{(1-\phi_j) X_{\mathrm{W_b}}} = \frac{\phi_j x_{\mathrm{G_b}}}{(1-\phi_j)(1-x_{\mathrm{G_b}})} = e^{-\Delta G_{\mathrm{ex}j}/(\mathrm{R}T)} \tag{3.13}$$

where ϕ_j is the fraction of first hydration shell sites in the protein j-state occupied by water molecules. We have introduced the molar fraction of cosolvent in the bulk binary solvent,

$$x_{\mathrm{G_b}} = \frac{X_{\mathrm{G_b}}}{X_{\mathrm{G_b}} + X_{\mathrm{W_b}}} \tag{3.14}$$

To note, if water is preferentially attached to the protein, $K_{\mathrm{ex}j} > 1$ ($\Delta G_{\mathrm{ex}j} < 0$), otherwise, when there is a preferential binding of protein with cosolvent molecules, $K_{\mathrm{ex}j} < 1$ ($\Delta G_{\mathrm{ex}j} > 0$). We assume that the exchange equilibrium processes are independent events, so that the probability that n water sites are occupied by water molecules and the remaining $m_j - n$ sites by cosolvent molecules is given by the binomial distribution, $p(n, m_j) = \frac{m_j!}{n!(m_j-n)!}\phi_j^n(1-\phi_j)^{m_j-n}$. Hence, by referring to Eq. 3.11, the molar concentration of the protein in the j-state dissolved in a binary solvent with all its m_j first hydration shell sites occupied by water molecules is given by $C_{\mathrm{P}_j(\mathrm{W_s}_j)_{m_j}} = C_j \phi_j^{m_j}$, where $C_j = \alpha_j^{-1} C_{\mathrm{P}}\, x_j = \sum_{n=0}^{m_j} C_{\mathrm{P}_j(\mathrm{W_s}_j)_n(\mathrm{G_s}_j)_{m_j-n}}$ is the total molar concentration of the protein in the j-state, independently on the occupation of the sites by water or cosolvent. Notice that $C_{\mathrm{P}} = c/M_1 = n/N_A$ is the nominal molar concentration of monomers in solution. As a consequence, in a binary solvent, the *effective* equilibrium constant K_{1j}, which describes the transformation of a protein molecule by the 1-state to the j-state, irrespective of the composition of the first hydration shell, and the related *effective* Gibbs free energy change are

$$K_{1j} = \frac{(\alpha_j^{-1} C_{\mathrm{P}}\, x_j)^{\alpha_j^{-1}}}{C_{\mathrm{P}}\, x_1} = K_{\mathrm{W}1j} \frac{\phi_1^{m_1} X_{\mathrm{W_b}}^{\alpha_j^{-1} m_j - m_1}}{\phi_j^{\alpha_j^{-1} m_j}} = e^{-\Delta G_{1j}/(\mathrm{R}T)} \tag{3.15}$$

The composition of the system is expressed by the nominal molar fractions of water, X_{W}, cosolvent, X_{G}, and protein monomers, X_{P}, with the straightforward condition $X_{\mathrm{W}} + X_{\mathrm{G}} + X_{\mathrm{P}} = 1$. Consequently, the nominal composition of the solvent is

$$x_{\mathrm{G}} = \frac{X_{\mathrm{G}}}{1 - X_{\mathrm{P}}} \tag{3.16}$$

Since X_{P} and x_{G} are fixed parameters characterizing the sample, in any conditions of protein distribution among the states and preferential solvation effects, the following two constraints should hold,

$$X_{\mathrm{W_b}} = (1 - X_{\mathrm{P}})(1 - x_{\mathrm{G}}) - X_{\mathrm{P}} \sum_{j=1}^{N_s} m_j x_j \alpha_j^{-1} \phi_j \tag{3.17}$$

$$X_{\mathrm{G_b}} = (1 - X_{\mathrm{P}}) x_{\mathrm{G}} - X_{\mathrm{P}} \sum_{j=1}^{N_s} m_j x_j \alpha_j^{-1} (1 - \phi_j) \tag{3.18}$$

Notice that the effective parameters K_{1j} and ΔG_{1j} can change with the composition (i.e. by varying X_{P} or x_{G}), whereas the exact thermodynamic parameters $K_{\mathrm{W}1j}$ and $\Delta G_{\mathrm{W}1j}$ as well as $K_{\mathrm{ex}j}$ and $\Delta G_{\mathrm{ex}j}$, which refer to the two elementary processes of Eqs. 3.10 and 3.12, should be independent on X_{P} and x_{G}. However, the Gibbs free energy change $\Delta G_{\mathrm{W}1j}$

can be affected by pH and ionic strength, which could be modified by the presence of cosolvent molecules, if they possess acid-base or ionic properties (such as for some of the modified-sugars exploited in this work, see Table 3.8). In order to deal with those cases, we separate an electrostatic term from all the other non-electrostatic terms [82, 6], $\Delta G_{\mathrm{W}1j} = \Delta G_{\mathrm{W,el},1j} + \Delta G_{\mathrm{W,nel},1j}$, and we write $\Delta G_{\mathrm{W,el},1j} = \alpha^{-1} G_{\mathrm{W,el},j} - G_{\mathrm{W,el},1}$ in the framework of the Debye-Hückel theory,

$$G_{\mathrm{W,el},j} = \frac{q_{\mathrm{e}}^2 Z_j^2}{8\pi\varepsilon_0\varepsilon R_j}\left(1 - \frac{\kappa_D R_j}{1 + \kappa_D(R_j + a)}\right) \tag{3.19}$$

In this equation, $q_{\mathrm{e}} = 1.6 \cdot 10^{-19}$ C is the charge of the proton, expressed in SI units, ε_0 is the vacuum permittivity, ε is the relative dielectric constant of the solvent, Z_j is the number of the elementary charges provided by the j-protein, which is assumed to be a spherical macroion with radius R_j, and a is the average radius of the all the microions (including protein counterions) in solution. To note, Z_j can be simply calculated as a function of pH considering the side chain pK_a values of the amino acids [46]. The reciprocal Debye-Hückel screening length, $\kappa_D = (2N_A q_{\mathrm{e}}^2 I/(\varepsilon_0\varepsilon k_B T))^{1/2}$ is another parameter of $G_{\mathrm{W,el},j}$ (k_B is Boltzmann's constant). It depends on the ionic strength due to the molar concentration C_i and the charge number z_i of all i-microions, $I = \frac{1}{2}\sum_i z_i^2 C_i$. On the basis of the electroneutrality condition, the molar concentration of protein counterions (assumed for the sake of simplicity to have a charge $|z_{\mathrm{ci}}| = 1$) should be $C_{\mathrm{ci}} = C_{\mathrm{P}} \sum_{j=1}^{N_s} x_j \alpha_j^{-1} |Z_j|$. We can hence write $I = I_S + I_{\mathrm{ci}}$, where I_S is the added ionic strength and I_{ci} is the one due to counterions. To note, I_S is calculated considering microions due to charged buffer molecules, if any, and microions provided by the cosolvent, in the case they are charged species. The non-electrostatic term $\Delta G_{\mathrm{W,nel},1j}$ includes all the other contributions to the thermodynamic stability of the j-protein state. Its temperature dependency, as well as the one of $\Delta G_{\mathrm{ex}j}$, is written according to classical thermodynamics, $\Delta G = \Delta G^{\circ} + (\Delta C_p - \Delta S^{\circ})(T - T_{\circ}) - \Delta C_p T \log(T/T_{\circ})$, where ΔG° and ΔS° are the changes of Gibbs free energy and entropy at the reference temperature $T_{\circ} = 298.15$ K, respectively, and ΔC_p is the change of the heat capacity at constant pressure, here considered to be independent on temperature. On the other hand, due to thermal expansion, also molar volumes are affected by temperature. Regarding water, according with Ref. [38], the molecular volume can be described by the approximation $\nu_{\mathrm{W_b}} = \nu^{\circ}_{\mathrm{W_b}} e^{\alpha_{\mathrm{W}}(T-T_{\circ}) + \frac{1}{2}\beta_{\mathrm{W}}(T-T_{\circ})^2}$, where the optimum values of the molar water volume at $T_{\circ}$, the thermal expansivity at $T_{\circ}$ and its first derivative are $\nu^{\circ}_{\mathrm{W_b}} = 0.018$ L, $\alpha_{\mathrm{w}} = 2.5 \cdot 10^{-4}$ K^{-1} and $\beta_{\mathrm{w}} = 9.8 \cdot 10^{-6}$ K^{-2}, respectively [84]. For cosolvent and protein molar volumes, we adopt a simpler approximation, just in terms of molar volumes and thermal expansivities at $T_{\circ}$: $\nu_{\mathrm{G_b}} = \nu^{\circ}_{\mathrm{G_b}} e^{\alpha_{\mathrm{G}}(T-T_{\circ})}$ and $\nu_{\mathrm{P}} = \nu^{\circ}_{\mathrm{P}} e^{\alpha_{\mathrm{P}}(T-T_{\circ})}$ The nominal molar concentration of monomeric proteins (seen in Eq. 3.15) is

$$C_{\mathrm{P}} = \frac{X_{\mathrm{P}}}{<\nu>} \tag{3.20}$$

$$\begin{aligned} <\nu> &= \nu_{\mathrm{P}} X_{\mathrm{P}} + (\nu_{\mathrm{W_b}}(1 - x_{\mathrm{G}}) + \nu_{\mathrm{G_b}} x_{\mathrm{G}})(1 - X_{\mathrm{P}}) \\ &\quad + X_{\mathrm{P}} \sum_{j=1}^{N_s} m_j x_j \alpha_j^{-1}((\nu_{\mathrm{W_{s}}_j} - \nu_{\mathrm{W_b}})\phi_j + (\nu_{\mathrm{G_{s}}_j} - \nu_{\mathrm{G_b}})(1 - \phi_j)) \end{aligned} \tag{3.21}$$

where the average molar volume $<\nu>$ is calculated as a function of the molar volume occupied by water and by cosolvent in the sites of the j-state of the protein, $\nu_{\mathrm{W_{s}}_j}$, and $\nu_{\mathrm{G_{s}}_j}$, respectively. In practice, only the former is considered to differ from the bulk value, since it has been widely demonstrated that hydration water has a more compact structure than bulk water [87, 77]. Accordingly, we write $\nu_{\mathrm{W}_{\mathrm{s}j}} = \nu_{\mathrm{W}_{\mathrm{b}}}/d_j$, where d_j is the relative mass density of hydration water, with typical values comprised in the range $1 \div 1.15$. By combining Eqs. 3.17, 3.18 and 3.14, it is straightforward to derive the cosolvent molar fraction of the bulk solvent as a function of both the fixed sample parameters, X_{P} and x_{G}, and the parameters depending on the interplay of the equilibrium processes, the molar fraction x_j of nominal protein monomers that are forming the j-state and the water occupation fraction ϕ_j of the first hydration shell of each j-state,

$$x_{\mathrm{G_b}} = \frac{x_{\mathrm{G}}(1 - X_{\mathrm{P}}) - X_{\mathrm{P}} \sum_{j=1}^{N_s} m_j x_j \alpha_j^{-1}(1 - \phi_j)}{1 - X_{\mathrm{P}}(1 + \sum_{j=1}^{N_s} m_j x_j \alpha_j^{-1})} \tag{3.22}$$

The nonlinear system of $2N_s$ equations, which includes Eq. 3.8, Eq. 3.13 (with $j = 1, N_s$) and Eq. 3.15 (with $j = 2, N_s$), in which the parameters $X_{\mathrm{W_b}}$, C_{P}, $<\nu>$ and $x_{\mathrm{G_b}}$ are obtained from Eqs. 3.17-3.22, respectively, contains the following $2N_s$ unknown variables: x_j and ϕ_j (both with $j = 1, N_s$).

The system is solved by a numerical iterative method as described in the Section 3.2.6.1 of the Supplementary Materials (Section 3.2.6). In such a way, we have a method able to derive, from the thermodynamic parameters ΔG_k°, ΔS_k° and ΔC_{p_k} that describe the two categories of elementary processes (non electrostatic contribution of protein state formation in water (Eq. 3.10) and water replacement of a cosolvent molecule over the surface of any protein state (Eq. 3.12)) the fraction x_j of nominal protein monomers distributed in the j-state and the fraction ϕ_j of the m_j first hydration shell sites over the protein surface occupied by water. Beside, we are able to calculate the cosolvent molar fraction of the bulk solvent, $x_{\mathrm{G_b}}$, the effective constants K_{1j} and the related Gibbs free energy change ΔG_{1j}. All these parameters are obtained as a function of the nominal protein molar fraction X_{P}, the nominal binary solvent composition x_{G}, the pH, the added ionic strength I_S and the temperature T.

3.2.3.2 Determination of SLDs

Results from this thermodynamic scheme allow also to calculate the SLDs of bulk solvent and protein hydration shells. As widely discussed by Ref. [83, 57], since the volume of the cosolvent molecule is much larger than the one of water, we have to consider that the cosolvent attached to the protein surface can in part occupy the hydration sites of the second hydration shell. As a consequence, preferential solvation effects will change the composition of a region in the vicinity of the protein surface, called local domain, which will encompass the hydration sites of both the first and the second shell. More in detail, the number of sites occupied by water and cosolvent in the first hydration shell (corresponding to the number of water and cosolvent molecules attached to the protein surface) are $N_{\mathrm{W},j,1} = m_j \phi_j$ and $N_{\mathrm{G},j,1} = m_j(1-\phi_j)$, respectively. Hence, the number of hydration sites of the second layer occupied by cosolvent molecules attached to the protein will be $k_j = m_j(1-\phi_j)(\nu_{\mathrm{G}_{\mathrm{s}j}} - \nu_{\mathrm{W}_{\mathrm{s}j}})/\nu_{\mathrm{W_b}}$. Indicating by $m_{j,2}$ the total number of hydration sites of the second layer, the ones that remain available to be occupied with the bulk solvent (with composition $x_{\mathrm{G_b}}$) will be $m_{j,2} - k_j$. We can then calculate the number of water and cosolvent molecules that occupies the available sites of the second hydration shell, according to $N_{\mathrm{W},j,2} = (m_{j,2} - k_j)(1 - x_{\mathrm{G_b}})/(1 + x_{\mathrm{G_b}}(\nu_{\mathrm{G_b}}/\nu_{\mathrm{W_b}} - 1))$ and

$N_{\mathrm{G},j,2} = (m_{j,2} - k_j) x_{\mathrm{G_b}} / (1 + x_{\mathrm{G_b}} (\nu_{\mathrm{G_b}} / \nu_{\mathrm{W_b}} - 1))$, respectively. On this basis, the cosolvent molar fraction of the local domain is

$$x_{\mathrm{G_{ld}}_j} = \frac{N_{\mathrm{G},j,1} + N_{\mathrm{G},j,2}}{N_{\mathrm{G},j,1} + N_{\mathrm{G},j,2} + N_{\mathrm{W},j,1} + N_{\mathrm{W},j,2}} \tag{3.23}$$

and the local domain molar volumes of water and cosolvent are

$$\nu_{\mathrm{W_{ld}}_j} = \frac{N_{\mathrm{W},j,1} \nu_{\mathrm{W_s}_j} + N_{\mathrm{W},j,2} \nu_{\mathrm{W_b}}}{N_{\mathrm{W},j,1} + N_{\mathrm{W},j,2}} \tag{3.24}$$

$$\nu_{\mathrm{G_{ld}}_j} = \frac{N_{\mathrm{G},j,1} \nu_{\mathrm{G_s}_j} + N_{\mathrm{G},j,2} \nu_{\mathrm{G_b}}}{N_{\mathrm{G},j,1} + N_{\mathrm{G},j,2}} \tag{3.25}$$

Hence, the SLDs of bulk solvent and local domain are

$$\rho_0 = \frac{x_{\mathrm{G_b}} b_{\mathrm{G}} + (1 - x_{\mathrm{G_b}}) b_{\mathrm{W}}}{x_{\mathrm{G_b}} \nu_{\mathrm{G_b}} + (1 - x_{\mathrm{G_b}}) \nu_{\mathrm{W_b}}} \tag{3.26}$$

$$\rho_{\mathrm{ld},j} = \frac{x_{\mathrm{G_{ld}}_j} b_{\mathrm{G}} + (1 - x_{\mathrm{G_{ld}}_j}) b_{\mathrm{W}}}{x_{\mathrm{G_{ld}}_j} \nu_{\mathrm{G_{ld}}_j} + (1 - x_{\mathrm{G_{ld}}_j}) \nu_{\mathrm{W_{ld}}_j}} \tag{3.27}$$

where $b_{\mathrm{W}} = r_{\mathrm{e}} N_{\mathrm{W,e}}$ and $b_{\mathrm{G}} = r_{\mathrm{e}} N_{\mathrm{G,e}}$ are the scattering lengths of water and cosolvent, $N_{\mathrm{W,e}}$ and $N_{\mathrm{G,e}}$ being the corresponding number of electrons and $r_{\mathrm{e}} = 0.28 \cdot 10^{-12}$ cm the classical radius of the electron. Considering the intrinsic low resolution of SAXS, also due to mobility effects over the protein surface, the calculation of both the form factors $P(q)$ (Eq. 3.7) and $P^{(1)}(q)$ (Eq. 3.9) with SASMOL is performed by assigning to all the sites of the first and the second hydration shell (their numbers are m_j and $m_{j,2}$, respectively) a unique SLD, corresponding to $\rho_{\mathrm{ld},j}$ (Eq. 3.27).

3.2.3.3 Effective protein-protein structure factor

The protein-protein structure factor $S(q)$ in the presence of a mixture of N_s protein states is due to a complex interplay of the partial structure factors $S_{j_1,j_2}(q)$ between any j_1, j_2 pair of states weighted by their relative populations, which in turn depend on pair interaction potentials $u_{j_1,j_2}(r)$. Here, according to Pedersen et al. [53], we adopt a simpler point of view by taking into account a unique *effective* radial interaction potential $u(r)$ between two protein particles, irrespective of their state. This potential is described by the HSDY (Hard-Sphere Double-Yukawian) model, $u(r) = u_{\mathrm{HS}}(r) + u_{\mathrm{YC}}(r) + u_{\mathrm{YA}}(r)$, which combines a hard-sphere (HS) term,

$$u_{\mathrm{HS}}(r) = \begin{cases} \infty & r < 2R \\ 0 & r > 2R \end{cases} , \tag{3.28}$$

and two Yukawian terms, described by the equation $u_{Yk}(r) = B_{1k}exp[-B_{2k}(r-2R)]/r$. They are a screened Coulombian (C) repulsive term, with $B_{1\mathrm{C}} = 4\pi Z^2 q_{\mathrm{e}}^2/(\varepsilon_0\varepsilon\,(1+\kappa_D R)^2)$ and $B_{2\mathrm{C}} = \kappa_D$, and an attractive (A) term, with $B_{1\mathrm{A}} = -2JR$ and $B_{2\mathrm{A}} = 1/d$. In these equations, R is the average protein radius. It is calculated as an average of the protein radii R_j of any state, according to $R = (1/{<}\alpha^{-1}{>})\sum_{j=1}^{N_s} x_j\alpha_j^{-1}R_j$, where ${<}\alpha^{-1}{>} = \sum_{j=1}^{N_s} x_j\alpha_j^{-1}$. The average net number of elementary electric charges is calculated in a similar manner, $Z = (1/{<}\alpha^{-1}{>})\sum_{j=1}^{N_s} x_j\alpha_j^{-1}Z_j$. The attractive term depends on two parameters, J, the energy when two proteins are at contact, ($r = 2R$), and the scale length d. All attractive contributions, such as van der Waals forces, dipole-dipole or similar interactions are represented by $u_{\mathrm{YA}}(r)$. In the presence of cosolvent, which can provide variations of the surface properties of proteins, the values of J and d can change in a way that is not easily rationalized. Therefore, we have decided to leave the two parameters free to change for each experimental condition investigated by SAXS. The calculus of $S(q)$ on the basis of $u(r)$ has been carried out by using the perturbation of the Percus-Yevick (PY) structure factor, $S_0(q)$, due to the two Yukawian terms, on the basis of the Random-Phase Approximation (RPA) [34, 91, 56]. Details are shown in Section 3.2.6.3 of the Supplementary Materials (Section 3.2.6).

3.2.3.4 Global-fit of SAXS data

On the basis of the model described in the previous sections, we are able to set a unique fit of a batch of N_c SAXS curves recorded for water solutions of the protein of interest (which can show different states) by varying protein concentration and in the presence of different amounts of a cosolvent. This so-called global-fit can include several series of SAXS measurements performed with distinct types of cosolvents, provided single samples never contain two or more types of cosolvents. More specifically, SAXS curves are labeled with $N_p = 4$ *curve parameters*: protein w/v concentration at $T_\circ$, c°, temperature, T, type of cosolvent, G, and its concentration at $T_\circ$, C_{G}°. The task is accomplished by minimising the merit function $\mathcal{H} = \chi^2 + \gamma L$, where χ^2 is the average reduced chi-square

$$\chi^2 = \frac{1}{N_c} \sum_{k=1}^{N_c} \frac{1}{N_{q,k}} \sum_{i=1}^{N_{q,k}} \left(\frac{\frac{d\Sigma}{d\Omega}_{k,\mathrm{expt}}(q_i) - \frac{d\Sigma}{d\Omega}_{k,\mathrm{theo}}(q_i)}{\sigma_k(q_i)} \right)^2 \tag{3.29}$$

In this equation, $\frac{d\Sigma}{d\Omega}_{k,\mathrm{expt}}(q_i)$ is k^{th} measured SAXS curve recorded over a number $N_{q,k}$ of q-points, $\frac{d\Sigma}{d\Omega}_{k,\mathrm{theo}}(q_i)$ is the theoretical curve calculated on the basis of Eq. 3.6 and $\sigma_k(q_i)$

is the experimental standard deviation. The other term of the merit function, L, is the regularization factor,

$$L = \sum_{i=1}^{2} \sum_{k=1}^{N_c} \sum_{p=1}^{N_p} \left(1 - \frac{X_{i,k'}}{X_{i,k}}\right)^2 , \tag{3.30}$$

which increases with the difference between the i^{th} single curve fitting parameter ($i = 1, 2$ refers to J and d, respectively) of the k-curve, $X_{i,k}$, and the one of the k'-curve, $X_{i,k'}$, where k' is the label of the curve having the same *curve parameters* of the k-curve but the p^{th} ($p = 1, N_p$ refers to c°, T, G and C°_{G}). The constant γ is selected in order to guarantee that when $\chi^2 \approx 1$, indicating a good fit, the product $\gamma\, L$ is $\approx 10\%$ of the merit function $\mathcal{H}$. The present model has been included in the freely available GENFIT software [81].

3.2.3.5 Myoglobin

According to a number of experimental as well as computational evidences [40, 55, 48, 54, 22, 5, 80, 20], myoglobin (MB) in solution at pH=5.0 and as a function of temperature can be present in three states, native (N), intermediate (I), and unfolded (U). The native state is monomeric and its form factor has been calculated on the basis of the PDB entry `1wla` [44]. The corresponding form factor has been then calculated with SASMOL. The average numbers of hydration sites in the first and in the second shell are found to be $m_{\mathrm{N}} = 404$ and $m_{\mathrm{N},2} = 465$, respectively. The intermediate state is considered to be a compact dimer [80], which has proven to maintain its active form [40, 48]. Its form factor has been calculated with SASMOL from the PDB entry `3vm9` [48]. The number of hydration sites are $m_{\mathrm{I}} = 753$ and $m_{\mathrm{I},2} = 827$. The unfolded state of MB has been described by a set of 50 conformations obtained by FOX, a home made software that preserves the secondary structure of a native structure and randomly modifies the Ramachandran angles of the residues that do not belong to helices or strands [45]. Steric clashes are avoided by controlling the overlap between the van der Waal spheres associated to each atom. The input PDB entry `1wla` has been adopted. The average form factor has been then calculated with SASMOL. The number of hydration sites in the first and in the second shell are $m_{\rm U} = 844$ and $m_{{\rm U},2} = 1141$, respectively. The form factors of the $N_s = 3$ states of MB are shown in the Fig. 3.32 of the Supplementary Materials (Section 3.2.6) in the form of semi-logarithmic and Kratky plots, together with the coupling function $\beta(q)$. The number of elementary charges for the N-state, Z_N, calculated on the basis of the primary sequence of MB and as a function of pH are reported in Table 3.8 of the Supplementary Materials (Section 3.2.6). For I and N states we have simply fixed $Z_I = 2Z_N$ and $Z_U = Z_N$.

3.2.3.6 Insulin

Insulin (IN) in water solution has been found to mainly form $N_s = 4$ folded states, corresponding to monomers (1), dimers (2), tetramers (4) and hexamers (6) [36, 29, 27, 89, 23, 96, 78, 49]. Monomers are formed by two polypeptide chains, named A and B, linked by two disulfide bridges. It is known that insulin is present in its hexameric form, which is the best way to store and stabilize the functional monomers. Once hexamers dissociate into monomers, dimers, and tetramers, they can be transported in the bloodstream and they are ready to exert their physiological activity [21]. The basic processes of oligomer formation have been identified as follows [36],

$$\begin{aligned} &2\,\mathrm{I}_1 \rightleftharpoons \mathrm{I}_2 \\ &2\,\mathrm{I}_2 \rightleftharpoons \mathrm{I}_4 \\ &\mathrm{I}_2 + \mathrm{I}_4 \rightleftharpoons \mathrm{I}_6 \end{aligned} \tag{3.31}$$

Related thermodynamic constants have been determined at room temperature. They are $\bar{K}_{12} = 2.22\ 10^5\ \mathrm{M}^{-1}$, $\bar{K}_{24} = 40\ \mathrm{M}^{-1}$ and $\bar{K}_{46} = 220\ \mathrm{M}^{-1}$ [36, 29]. These constants are connected to the effective constants defined in Eq. 3.15 by the following relationships: $\bar{K}_{12} = K_{12}^2$, $\bar{K}_{24} = (K_{14}/K_{12})^4$ and $\bar{K}_{46} = K_{16}^6/(K_{14}^4 K_{12}^2)$. Form factors of the different states have been calculated with SASMOL on the basis of the PDB entry `3aiy` [51]. For the monomer, only chains A and B have been considered, for the dimer the chains A-D, for the tetramer the chains A-H and for the hexamer the whole PDB file (A-L chains). Fig. 3.33 of the Supplementary Materials (Section 3.2.6) reports semi-logarithmic and Kratky plots of the form factors of the $N_s = 4$ states and their coupling function $\beta(q)$. For the monomer, the average numbers of hydration sites in the first and in the second shell are found to be $m_1 = 199$ and $m_{1,2} = 268$, respectively. For the dimer, the tetramer and the hexamer corresponding values are: $m_2 = 320$, $m_{2,2} = 385$; $m_4 = 548$, $m_{4,2} = 625$; $m_6 = 741$, $m_{6,2} = 739$. In Table 3.8 of the Supplementary Materials (Section 3.2.6) the number of elementary charges for the 1-state, Z_1, which has been obtained considering the primary sequence of IN and the pH of the solution are reported. For the other states we have simply fixed $Z_j = \alpha_j Z_1$.

3.2.4 Modified sugars results

3.2.4.1 Myoglobin

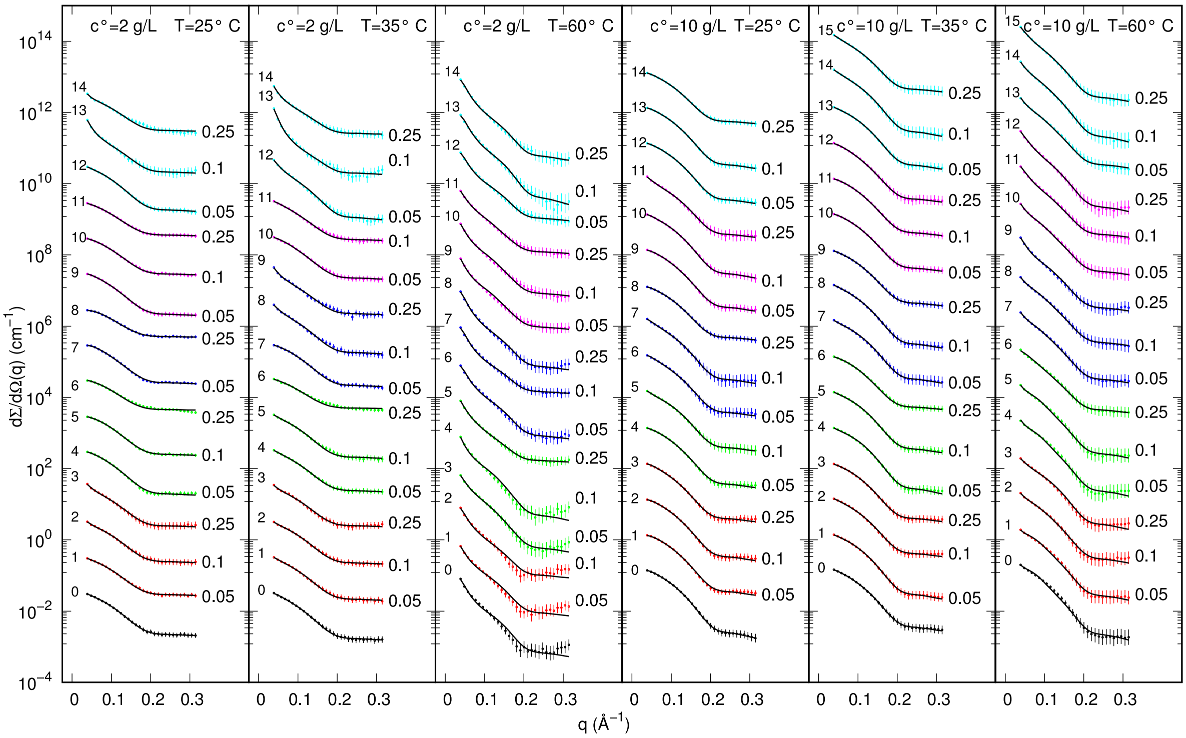


FIGURE 3.26: Experimental SAXS curves of MB in 10 mM phosphate buffer (pH=5) with and without ExtremoChem modified-sugar superimposed with the best fits obtained with GENFIT (solid lines). Colors refer to the following conditions: no-modified-sugar (black), EC312 (red), EC101 (green), EC311 (blue), EC202 (magenta), EC212 (cyan). Whenever present, the modified-sugar concentration is reported on the right side of each curve in molar unit. Each column refers to a fixed temperature and MB concentration, as indicated on the top. Curves are multiplied by the factor 10^k, k being reported on the top right of each curve. Experimental standard deviations are reported as error bars every 10 points, for clarity.

SAXS curves recorded at the Elettra synchrotron (Austrian SAXS beam-line) for samples of myoglobin in the presence of five different ExtremoChem modified-sugars, by varying protein or modified-sugar concentration as well as temperature, are shown in Fig. 3.26 in the form of semi-logarithm plots. To note, several curves show an upward curvature at low q, suggesting a predominant long-range attraction among the protein particles. Kratky plots, shown in Fig. 3.34 of the Supplementary Materials (Section 3.2.6), reveal, on the one hand, the presence of a main peak at a q-position that changes as a function of sample composition and temperature and, on the other hand, the absence of an asymptotic behavior at high q. These features suggest that the aggregation number of myoglobin can change with sample compositions and that most of the protein states are compact, also at the highest temperatures. Hence, since the information content of the SAXS dataset on the U-state is low, the ensemble of unfolded conformations calculated with the FOX method [45] has been left fixed. On the basis of these preliminary observations, the whole set of $N_c = 92$ SAXS curves has been globally analyzed by using the new method introduced in Section 3.2.3. Three possible myoglobin states have been taken into account: two of them, the native monomer (N) and the intermediate dimer (I), are biologically active states, whereas the monomeric unfolded (U) state represents a denatured (inactive) form of the protein. The full list of the model parameters, together with their short description and the validity range we have defined is reported in Table 3.9 of the Supplementary Materials (Section 3.2.6). Best fitting curves are reported as solid lines in Fig. 3.26: the high quality of the fit, in the entire q-range, can be appreciated. By fixing the dimensionless regularization parameter $\gamma = 10^{-7}$, the overall merit function $H = 1.15$ has been obtained, corresponding to $\chi^2 = 1.03$ ($\gamma L = 0.12$, see Eq. 3.30). The thermodynamic fitting parameters obtained by the simultaneous analysis of the whole set of SAXS data are reported in Table 3.6. First of all, results indicate that, in water at pH=5, the NI transition of myoglobin from native monomer to intermediate dimer has a low non-electrostatic reference Gibbs free energy barrier ($\Delta G^{\circ}_{\mathrm{W,nel},NI} = (2.95 \pm 0.03)\ \mathrm{kJ\,mol^{-1}}$,

j_1j_2	$\Delta G^\circ_{\mathrm{W,nel},j_1j_2}$ kJ mol^{-1}	$\Delta S^\circ_{\mathrm{W}j_1j_2}$ J mol^{-1} K^{-1}	$\Delta C_{p\mathrm{W}j_1j_2}$ J mol^{-1} K^{-1}
NI	2.95±0.03	564±6	0±2
NU	167±3	1600±500	8400±400
j	$\Delta G^\circ_{\mathrm{ex}j}$ kJ mol^{-1}	$\Delta S^\circ_{\mathrm{ex}j}$ J mol^{-1} K^{-1}	$\Delta C_{p\mathrm{ex}j}$ J mol^{-1} K^{-1}
		EC312	
N	1.5±0.2	−10±10	−7±3
I	−8±1	10±10	−9±6
U	−8.7±0.5	−10±9	2±3
		EC101	
N	2.2±0.1	−9±4	3±3
I	−8.5±0.7	9±6	4±7
U	−4±1	−10±7	0±8
		EC311	
N	−0.4±0.6	2±4	1±5
I	−2.4±0.8	9±9	3±7
U	−3±4	−9±2	−1±4
		EC202	
N	−3±3	−10±10	0±4
I	−8.5±0.5	1±4	7±6
U	−8±1	6±9	−7±7
		EC212	
N	−1.3±0.4	9±7	−7±8
I	−7±2	−8±8	3±3
U	−9.3±0.5	3±3	−2±4

TABLE 3.6: Thermodynamic fitting parameters obtained by the global-fit of MB SAXS curves shown in Fig. 3.26. $\Delta G^\circ_{\mathrm{W,nel},j_1j_2}$, $\Delta S^\circ_{\mathrm{W}j_1j_2}$ and $\Delta C_{p\mathrm{W}j_1j_2}$: changes of non-electrostatic reference Gibbs free energy, reference entropy and heat capacity at constant pressure, respectively, occurring at the j_1j_2 transition; $\Delta G^\circ_{\mathrm{ex}j}$, $\Delta S^\circ_{\mathrm{ex}j}$ and $\Delta C_{p\mathrm{ex}j}$: changes of reference Gibbs free energy, reference entropy and heat capacity at constant pressure, respectively, occurring at the modified-sugar-water exchange over the j-state.

Table 3.6) when compared to the value related to the NU transition from native monomer to unfolded monomers ($\Delta G^\circ_{\mathrm{W,nel},NU} = (167 \pm 3)$ kJ mol^{-1}). Another fitting parameter, which is relevant in determining the effect of temperature, is the reference entropy variation, $\Delta S^\circ_{\mathrm{W}j_1j_2}$, related to the two NI and NU processes. The NI transition from native monomer to intermediate and still active dimer causes an increase of reference entropy, an effect that can be explained considering that this process determines the release of hydration water to the bulk solution. Indeed, since there are 404 hydration water molecules in the monomeric N-state and 753 in the dimeric I-state (see Sec. 3.2.3.5), 28 molecules of water for each monomer are released in solution when the dimer is formed, resulting in an increase of the reference entropy up to (564 ± 6) J mol^{-1} K^{-1} (Table 3.6). On the other hand, when monomeric myoglobin switches from native to unfolded state (NU transition), a different scenario emerges. Although in the U-state the protein shell is surrounded by 844 water molecules, a value much higher if compared to 404 molecules that encircle the N-state, with a concomitant decrease of entropy, the formation of an unfolded disordered state is surely accompanied by an huge increase of entropy, so that the balance between the two phenomena leads to the observed large and positive value of the reference entropy change $\Delta S^{\circ}_{\mathrm{W}NU} = (1600 \pm 500)$ J mol^{-1} K^{-1}, Table 3.6. The almost zero variation of the $\Delta C_{p_{\mathrm{W}NI}}$ and the large and positive value $\Delta C_{p_{\mathrm{W}NU}} = (8400 \pm 400)$ J mol^{-1} K^{-1} are expected, considering the large accessible surface area of the protein unfolded state [67]. The other fitting parameters reported in Table 3.6 regard the changes of reference Gibbs free energy, reference entropy and heat capacity at constant pressure that occur when a modified-sugar molecule bound to the myoglobin surface in each of the three envisaged j-states (N, I and U) is replaced by a water molecule. In general, we observe that whereas the experimental uncertainty of $\Delta G^{\circ}_{\mathrm{ex}j}$ is low (on average in the order of 1%), the ones of $\Delta S^{\circ}_{\mathrm{ex}j}$ and $\Delta C_{p_{\mathrm{ex}j}}$ are much larger, a results that can be in part explained considering that we have investigated our samples only at three different temperatures. These high uncertainties reflect the correctness of the global-fit SAXS data analysis method (see details in Sec. 3.2.6.2 of the Supplementary Materials (Section 3.2.6)), which does not lead to an overestimation of the parameters when their information content in the dataset is low.

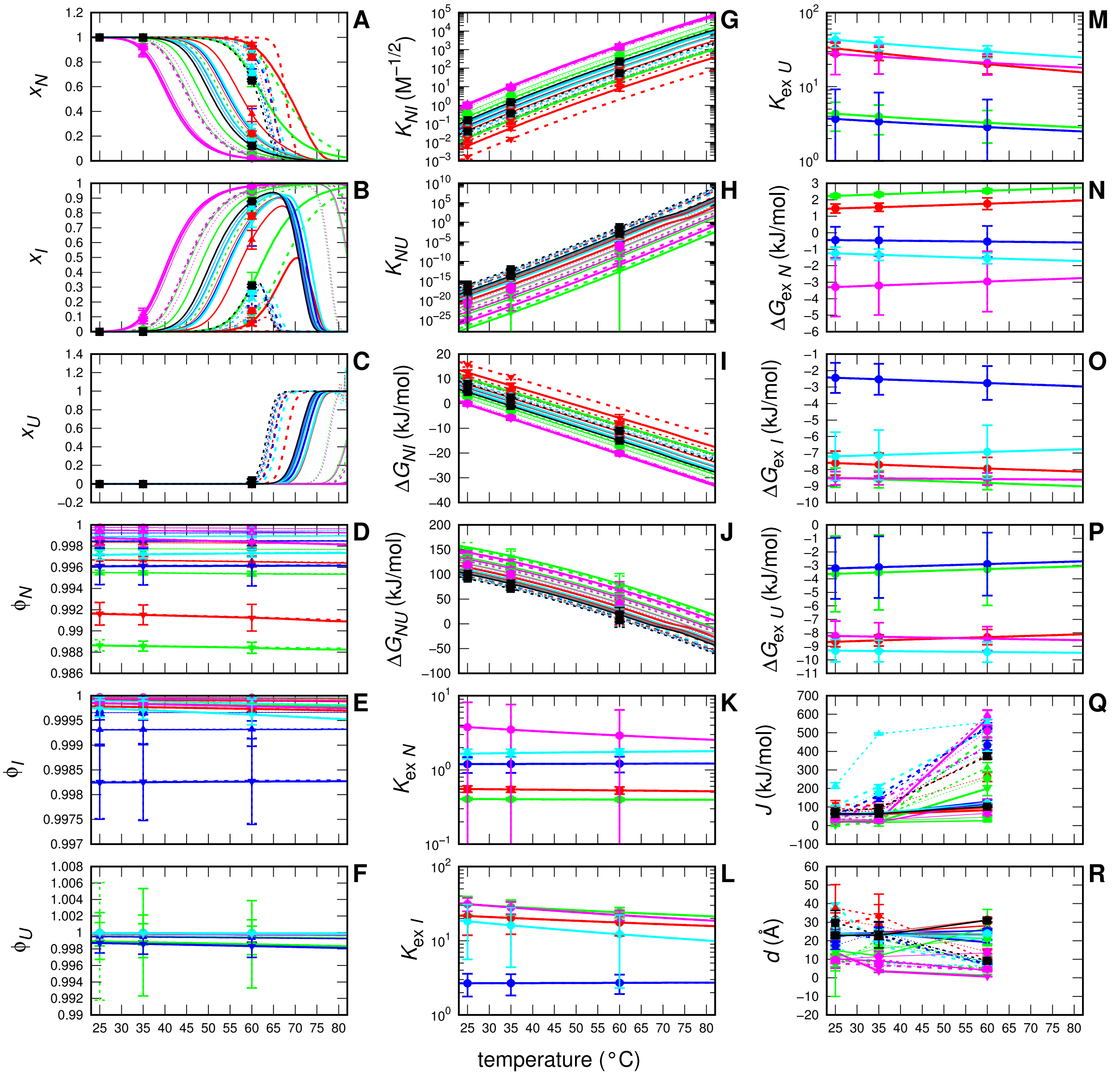


FIGURE 3.27: Temperature behaviors of the most relevant physical-chemical parameters obtained by the global-fit of MB SAXS curves shown in Fig. 3.26. Color refers to: no-modified-sugar (black), EC312 (red), EC101 (green), EC311 (blue), EC202 (magenta), EC212 (cyan). Thickness refers to: 0.05 M (thin), 0.10 M (intermediate), 0.25 M (thick). Point-type refers to: no-modified-sugar (square), 0.05 M (circle), 0.10 M (up-sided triangle), 0.25 M (down-sided triangle). Dotted and solid lines refers to $c^\circ = 2$ g/L and $c^\circ = 10$ g/L, respectivley.

The meaning of all the fitting parameters shown in Table 3.6 can be better appreciated by considering the temperature dependency of the most relevant physical-chemical parameters inherent to the adopted model that have been derived by them. Their trends are shown in Fig. 3.27. Notice that in this figure the colors and the thickness (together with the symbols) of the curves have been assigned according to the type and the concentration of the modified-sugar, respectively, whereas dotted and solid lines refer to 2 or 10 g/L myoglobin concentration, respectivley. Panels **A**, **B** and **C** report the molar fractions of the nominal myoglobin monomers distributed into the three different N, I or U states (x_N, x_I and x_U, respectivley). It is possible to appreciate that, even if at 25° C the N monomers are the only fraction present in solution, the temperature rise gradually affects the protein state, resulting in the decay of the monomeric N-state and the concomitant formation of dimeric I-state. It is known that the oxygen binding rate constant of myoglobin dimer is similar to that of the monomer, whereas the oxygen dissociation rate constant of the dimer is smaller than that of the monomer [48]. Hence, our results could provide suggestions concerning monomer-dimer function and role. However, the particular pH and buffer conditions which do not resemble *in vivo* conditions, suggest not to infer them by this experimental set-up. To note, although in our experiment we did not reach temperatures higher than 60° C, the adopted model with the fitting parameters derived by the set of SAXS data allows to predict that at higher temperatures the population of the U-state grows at the expense of the I-state. The fractions ϕ_j of first hydration shell sites of the j protein state occupied by water are reported in panels **D**, **E** and **F**. Since in our samples the presence of water is dominant, values of ϕ_j are very close to 1, with small but detectable differences, depending on the modified-sugar type. Such small differences, on the basis of Eq. 3.15, are sufficient to describe the modified-sugar-induced modification of the effective equilibrium constant K_{Nj} describing transition from the N-state to the j-state ($j = I, U$): results are shown in Fig. 3.27, panels **G** and **H**. Corresponding Gibbs free energy changes ΔG_{Nj}, which comprise both the electrostatic and the non-electrostatic contributions, are shown as a function of T in panels **I** and **J**. Exchange modified-sugar-water equilibrium constants $K_{\mathrm{ex}j}$ for each of the three j-states are reported in panels **K**, **L** and **M** and corresponding Gibbs free energy changes $\Delta G_{\mathrm{ex}j}$ are in panels **N**, **O** and **P**. Notice that both parameters do not depend on protein or modified-sugar concentration, but only on modified-sugar type. Finally, in panels **Q** and **R**, we report the depth of the attraction protein-protein potential J and its scale length d, which have been treated as single-curve fitting parameters.

Protein-protein structure factors $S(q)$ (Eq. 3.32 of the Supplementary Materials (Section 3.2.6)), calculated with the fitting parameters and included in the fitted $\frac{d\Sigma}{d\Omega}_{k,\mathrm{expt}}(q_i)$ function (Eq. 3.29), are plotted in Fig. 3.35 of the Supplementary Materials (Section 3.2.6). Corresponding effective radial interaction potentials $u(r)$ are reported in Fig. 3.36 of the Supplementary Materials (Section 3.2.6).

Other fitting parameters of the model are the relative densities of hydration water, d_j. We have found similar values for each of the $N_s = 3$ MB states, with an average value of 1.07 ± 0.02. The average radius R_j of N, I and U states found by the global-fit are (17.0 ± 0.2) Å, (26.7 ± 0.3) Å and (43 ± 2) Å, respectively. The ionic strength due to the buffer results (10.0 ± 0.1) mM.

We discuss in the next paragraphs results obtained in the absence of modified-sugars and in the presence of each of the five investigated modified-sugars.

3.2.4.2 Myoglobin without and with modified-sugar

In the absence of modified-sugars, MB at 2 g/L maintains its monomeric form (N-state) up to about 60° C (Fig. 3.27, panel **A**, black dotted lines), whereas the molar fraction of nominal MB monomers that are forming dimers (I-states) reaches the maximum peak of $x_I \approx 0.3$ (Fig. 3.27, panel **B**, black dotted lines). At higher temperatures, dimers sharply disappear and the unfolded U state becomes the predominant species in solution (Fig. 3.27, panel **C**, black dotted lines). On the contrary, at 10 g/L MB starts its transition from monomer to dimer at 45° C (Fig. 3.27, panel **A**, black solid lines) and up to 65° C the molar fraction of nominal MB monomers that are forming dimers is as large as $x_I \approx 0.9$ (Fig. 3.27, panel **B**, black lines).

The addition of the modified-sugars produces different effects depending on the type of the compound used, but it is in general evident that when the cosolvents are used together with the highest concentration of protein, MB tends to have a marked transition from monomer to dimer and it becomes unfolded at temperature higher than 70° C. On the other hand, MB at 2 g/L shows a different behavior, leaving out the dimeric form, except for two cosolvents that form ionic species in solution (EC101 and EC202).

3.2.4.3 Myoglobin with EC312

Myoglobin at 2 g/L in the presence of EC312 maintains its native monomeric state with a slow transition to dimers at ≈ 60° C, which slightly depends on EC312 concentration (Fig. 3.27, panel **A**, red lines). Dimers (I-state) do not overcome the fraction $x_I \approx 0.2$ of the myoglobin molecules in solution and gradually decrease and disappear at 65° C with the development of the unfolded state (Fig. 3.27, panel **B**, red lines). On the contrary, myoglobin at 10 g/L in the presence of EC312 is prone to form dimers at 50° C when the EC312 concentration is 0.05 M, leading to a solution rich in dimers ($x_I \approx 0.9$) that unfold at 75° C. At increasing concentration of EC312, the NI transition shifts from 55° C to 70° C, twenty degrees more than what occurs to the protein without modified-sugar. These results are also described by the behavior of the effective equilibrium constant K_{NI} (Fig. 3.27, panel **G**, red curves). For both 2 and 10 g/L MB concentration (dotted and solid red curves), K_{NI}, at increasing concentration of EC312, is lower than the value without EC312 (black lines). This aspect underlines the tendency of MB with EC312 to maintain its monomeric N-state for temperatures higher than the protein without EC312. Moreover, the temperature increase leads to an increase of K_{NI}, corresponding to a a preference for the dimeric state. For MB in the monomeric N-state, the exchange constant $K_{\text{ex}N}$ owns values lower than 1 (Fig. 3.27, panel **K**, red line), indicating a preference to be surrounded by EC312. On the other hand, the dimeric and the unfolded states show an increase of $K_{\text{ex}j}$ (panels **L** ($j = I$) and M ($j = U$)), suggesting the preference of these MB states to be solvated by water. The effect of EC312 in modifying protein-protein long range interactions is not marked, as can be observed by comparing the structure factors

$S(q)$ shown in Fig. 3.35 of the Supplementary Materials (Section 3.2.6) (red and black curves) and the corresponding $u(r)$ reported in Fig. 3.36 of the Supplementary Materials (Section 3.2.6). To note, at 60° C and 2 g/L myoglobin, a condition close to the NU transition, stronger attractive interactions among proteins have been seen, both with and without EC312, whereas at 60° C and 10 g/L, when most of the proteins are I-dimers, a less marked attraction is seen.

3.2.4.4 Myoglobin with EC101

Depending on its concentration, EC101 strongly affects the transition of MB from native monomer to intermediate dimer (Fig. 3.27, panels **A** and **B**, green lines). While, at lower modified-sugar concentration, the decay of N-monomers in favor of I-dimers begins $\approx$ 15° C earlier than for the samples without EC101 at both 2 or 10 g/L myoglobin, by increasing the EC101 concentration this transition occurs at higher temperatures. In particular, dimers begin to be present in solution at $\approx$ 40° C and subsequently totally substitute the N-monomers. The unfolded state is not present, except at temperatures above 80° C and with lower concentration of modified-sugar. The trends of the effective constant K_{NI} (Fig. 3.27, panel **G**) also confirms that by increasing EC101 concentration, especially at 10 g/L, the protein tends to remain in the monomeric N-state at higher temperatures than in the absence of EC101. Concerning the unfolded state, the very low values of K_{NU} (Fig. 3.27, panel **H**) show that there is no propensity for the protein to unfold except for temperatures higher than $\approx$ 80° C and in presence of the lowest EC101 concentration. The exchange constant $K_{\mathrm{ex}j}$ varies according to the type of protein state.

When MB is the N state, $K_{\mathrm{ex}N}$ is less than one (Fig. 3.27, panel **K**, green line), showing its preference to be surrounded by modified-sugar, while for the intermediate and the unfolded states (panels **L** and **M**, green lines), $K_{\mathrm{ex}j}$ is greater than one, underlining the preference of the protein in such states to be surrounded by water. Likewise the EC312 case, also EC101 shows weak effects in modifying protein-protein long range interactions (Figs. 3.35 and 3.36 of the Supplementary Materials (Section 3.2.6), green and black curves), confirming the presence of more marked attractions at 60° C and 2 g/L myoglobin, which are weaker at 60° C and 10 g/L.

3.2.4.5 Myoglobin with EC311

Myoglobin, at 2 and 10 g/L, in the presence of EC311, retains its monomeric N-state up to 55° C and 45° C, respectively (Fig. 3.27, panels **A** and **B**, blue lines), similarly to the protein in the absence of EC311, showing only a slight dependence on the EC311 concentration. At 2 g/L and in the presence of EC311, myoglobin appears to be present mainly in the form of N-monomer, except for a small gap between 55° C and 70° C, in which a small amount of dimer starts to grow, but it does not exceed the fraction ≈ 0.3 of the particles in solution. At 70° C all dimers formed with myoglobin 2 g/L are unfolded, while, at MB 10 g/L, I-dimers' fraction reach ≈ 0.9 and then disappear, with an increment of the unfolded state at 75° C. The equilibrium constant K_{NI} (Fig. 3.27, panel **G**, blue curves) slightly depends on EC311 concentration, which in turn resembles the one of protein in absence of EC311 (black lines). Results indicate that, with increasing quantities of EC311, the value of K_{NI} decreases, highlighting a tendency of the protein to be present in its monomeric N-state at higher temperatures in respect to the protein in the absence of EC311. A similar behavior is found also during the transition NU: the low K_{NU} values (Fig. 3.27, panel **H**, blue curves) confirm the propensity of the protein, at low temperatures, to be present in the N-state until 60° C. The exchange constant $K_{\mathrm{ex}j}$ is close to 1 when the protein is present in the N-state (Fig. 3.27, panel **K**, blue curve), suggesting that there is no preference to be surrounded by water or by EC311. On the other hand, when we consider the intermediate and the unfolded state, the $K_{\mathrm{ex}j}$ values rise slightly (Fig. 3.27, panels **L** and **M**, blue curves), suggesting a preference of MB in these states to be surrounded by water. Also for EC311, more pronounced effects on protein-protein long range attractions (Figs. 3.35 and S5 of the Supplementary Materials (Section 3.2.6), blue curves) are seen at 60° C and 2 g/L myoglobin and moderate effects are seen both at 60° C and 10 g/L and at 35° C and 2 g/L.

3.2.4.6 Myoglobin with EC202

In the presence of EC202, the behavior of MB at 2 and 10 g/L is quite similar (Fig. 3.27, panels **A** and **B**, magenta lines), showing a bigger shift if compared to the protein without EC202, which increases additionally as a function of the EC202 concentration. This means that MB switches from N-monomer to dimer at lower temperatures with respect to the protein without EC202 (black lines). The transition occurs at around 40° C for MB 10 g/L, ten degrees before the normal transition temperatures of the protein without EC202. A bigger effect is evident for MB 2 g/L, when the protein, in presence of EC202, has the NI transition that occurs at 45° C, twenty degrees before the protein without modified-sugar in solution. The unfolded fraction is almost absent, with a slight onset at the lowest modified-sugar concentrations at around 80° C. The trends of K_{NI} are almost independent on MB concentration (Fig. 3.27, panel **G**, magenta lines) and only slightly dependent on EC202 concentration. Results confirm the tendency of the protein to be in the intermediate state at lower temperatures compared to what happens in absence of EC202. The transition from native to unfolded state, on the other hand, is disadvantaged as the K_{NU} value is almost constantly lower than 1, except for temperatures higher than 80° C (Fig. 3.27, panel **H**, magenta lines). Like for the EC312 case, also with EC202 there is a slight dependence on the modified-sugar concentration, without any effect due to protein concentration. Indeed, curves of MB 2 and 10 g/L are almost superimposed. $K_{\mathrm{ex}j}$, which indicates the protein preference to be surrounded by water or modified-sugar, is much greater than 1 in each of the three envisaged states. In particular, a slight decreasing trend of the $K_{\mathrm{ex}j}$ parameter can be noted as a function of temperature, which, however, is not considered very relevant. The modified-sugar EC202 shows, in general, week protein-protein long range attractions (Figs. 3.35 and S5 of the Supplementary Materials (Section 3.2.6), magenta curves), the most relevant occurring at 60° C both at 2 and 10 g/L myoglobin.

3.2.4.7 Myoglobin with EC212

Myoglobin at 2 or 10 g/L, in the presence of EC212, retains the N state up to 55° C and 45° C, respectively, showing only a slight dependence on the EC212 concentration (Fig. 3.27, panel **A** and **B**, cyan curves). While MB at 2 g/L does not show a fraction of dimers greater than $x_I \approx 0.3$, at 10 g/L MB dimers' fraction reach ≈ 0.9. In the first case (2 g/L MB), at around 70° C the dimer has completely disappeared, replaced by the unfolded state, while at 10 g/L, a similar behavior happens at temperatures above 75° C. Both x_N and x_I curves follow a trend very similar to that of protein in absence of modified-sugars (black lines). The effective equilibrium constants K_{NI} (Fig. 3.27, panel **G**, cyan curves) are almost overlapping to the values in absence of EC212 (black lines). The exchange constant $K_{\mathrm{ex}j}$ is always greater than 1 for each of the considered states (N, I and U), highlighting a constant preference of the protein to be surrounded by water molecules.

3.2.4.8 Insulin

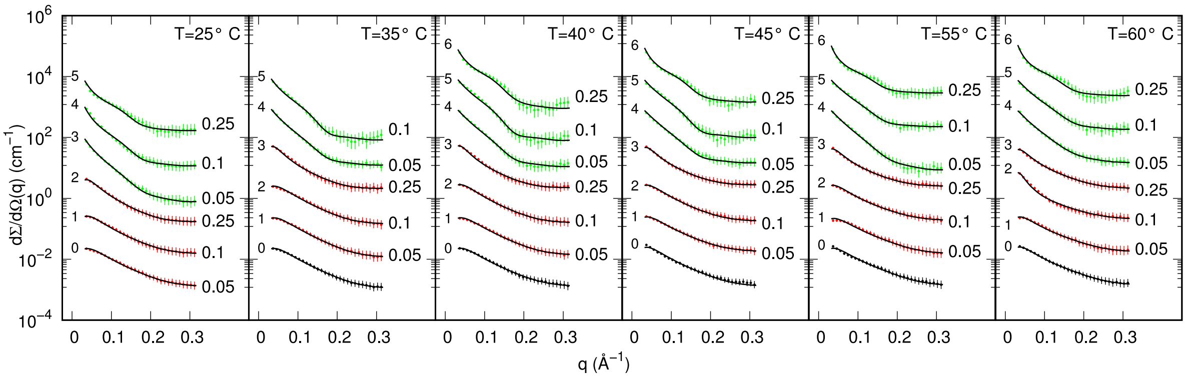


FIGURE 3.28: Experimental SAXS curves of 2 g/L IN in 10 mM phosphate buffer (pH=3) with and without ExtremoChem modified-sugar superimposed with the best fits obtained with GENFIT (solid lines). Colors refer to the following conditions: no-modified-sugar (black), EC312 (red), EC101 (green). Whenever present, the modified-sugar concentration is reported on the right side of each curve in molar unit. Each column refers to a fixed temperature, as indicated on the top. Curves are multiplied by the factor 10^k, k being reported on the top right of each curve.. Experimental standard deviations are reported as error bars every 10 points, for clarity.

SAXS curves recorded as a function of temperature for 2 g/L insulin in the presence of two modified-sugars, EC312 and EC101, are shown in form of semi-logarithm plots in Fig. 3.28. We first observe that several curves show an upward curvature at low q, indicating the prevalence of attraction forces at long range among the particles. In Fig. 3.37 of the Supplementary Materials (Section 3.2.6), Kratky plots of the experimental SAXS curves are shown: in all cases the presence of a main peak and the absence of asymptotic trends at high q revels the presence of compact IN shapes, with possible different aggregation states. On this basis, the simultaneous analysis of the $N_c = 40$ SAXS curves shown in Fig. 3.28 with the model introduced in Sec. 3.2.3 has been carried out by considering four possible states: monomer (1), dimer (2), tetramer (4), and hexamer (6). Table 3.10 of the Supplementary Materials (Section 3.2.6) reports the complete list of the model parameters, their description and the validity range.

Fitting curves, shown as solid lines in Fig. 3.28, are well superimposed to the experimental curves in the entire q-range. The regularization parameter γ has been fixed to 10^{-7}, leading to a merit function $H = 0.825$ and a corresponding $\chi^2 = 0.764$ ($\gamma L = 0.061$, Eq. 3.30). The main thermodynamic fitting parameters are reported in Table 3.7. We first observe that the non-electrostatic contribution of the reference Gibbs free energy changes, $\Delta \bar{G}^{\circ}_{\mathrm{W,nel},j_1 j_2}$, related to the three processes shown in scheme 3.31, occurring in water at pH=3, are always negative, suggesting the presence of mechanisms other than charge-charge interactions that favor the formation of IN oligomers. We also notice that the reference entropy changes related to the three processes are positive, a results that can be understood considering the release of water molecules in the bulk when these oligomers

j_1j_2	$\Delta\bar{G}^\circ_{\mathrm{W,nel},j_1j_2}$ kJ mol^{-1}	$\Delta\bar{S}^\circ_{\mathrm{W}j_1j_2}$ J mol^{-1} K^{-1}	$\Delta\bar{C}_{p\mathrm{W}j_1j_2}$ J mol^{-1} K^{-1}
12	−25.4±0.3	302±3	−5200±700
24	−16±1	60±10	−100±800
46	−28.5±0.3	396±4	−6400±700
j	$\Delta G^\circ_{\mathrm{ex}j}$ kJ mol^{-1}	$\Delta S^\circ_{\mathrm{ex}j}$ J mol^{-1} K^{-1}	$\Delta C_{p\mathrm{ex}j}$ J mol^{-1} K^{-1}
		EC312	
1	−9.3±0.8	2±5	0±4
2	1.5±0.3	10.0±0.1	3±5
4	−8±2	8±5	−5±6
6	−9.3±0.1	−1±5	−2±4
		EC101	
1	−9.7±0.1	6±7	8±9
2	−9.5±0.4	4±9	−9.5±0.2
4	4.3±0.2	−5±2	−10±2
6	3.3±0.4	−8±2	−9±6

TABLE 3.7: Thermodynamic fitting parameters obtained by the global-fit of IN SAXS curves shown in Fig. 3.28. $\Delta\bar{G}^\circ_{\mathrm{W,nel},j_1j_2}$, $\Delta\bar{S}^\circ_{\mathrm{W}j_1j_2}$ and $\Delta\bar{C}_{p\mathrm{W}j_1j_2}$: changes of non-electrostatic reference Gibbs free energy, reference entropy and heat capacity at constant pressure, respectively, occurring at the j_1j_2 transition (Eq. 3.31); $\Delta G^\circ_{\mathrm{ex}j}$, $\Delta S^\circ_{\mathrm{ex}j}$ and $\Delta C_{p\mathrm{ex}j}$: changes of reference Gibbs free energy, reference entropy and heat capacity at constant pressure, respectively, occurring at the modified-sugar-water exchange over the j-state.

are formed. Indeed, according to the number of waters sites found by SASMOL in the first hydration shell of the four species (Section 3.2.3.6), the numbers of water that are released due to the formation of dimers, tetramers or hexamers are are 78, 92, and 127, respectively. The heat capacities at constant pressure are found to be negative (Table 3.7) and affected by a quite large uncertainty (≈ 750 J mol^{-1} K^{-1}). According to Ref. [95], negative values of heat capacity change are due to the fragility of hydrogen bonds between water molecules at the hydrophobic interfaces. However, of all the major thermodynamic variables measured for proteins, heat capacity is the one with the most different set of definitions and the richest set of implications for protein folding and binding. Its sign can distinguish apolar from polar solvation, and it imparts a temperature dependence to entropy and enthalpy that may change their signs and determine which of them will dominate [64]. The other thermodynamic parameters shown in Table 3.7 regards the modified-sugar-water exchange in the surface of the four states of insulin that can be found in solution. The reference Gibbs free energy changes are obtained with low standard deviations (in the order of few percent), whereas larger uncertainties have been found for the reference entropy and heat capacity at constant pressure changes, confirming, such as for the MB case, that only a rough estimation of them can be derived from the SAXS dataset.

To fully understand the meaning of the fitting results, we report in Fig. 3.29 the temperature behavior of all the physical-chemical parameters of the model derived by the fitting parameters. To note, black curves refer to samples without modified-sugar, whereas red and green curves are devoted to EC312 and EC101 compounds, respectively. In detail, panel **A**-**D** show the trends of the four fractions x_1, x_2, x_4 and x_6, respectivley. Panels **E**-**H** reports the fraction ϕ_j of first hydration shell occupied by water in the j-state ($j = 1, 2, 4, 6$). Effective equilibrium constants of the three processes reported in scheme 3.31 are shown in panels **I**-**K** and corresponding Gibbs free energy changes (including both non-electrostatic and electrostatic terms) in panels **L**-**N**. Regarding the modified-sugar-water exchange processes, equilibrium constants and Gibbs free energy changes are reported in panels **Q**-**T** and **U**-**X**, respectively. Finally, panels **O** and **P** show the trend of the depth J and the scale length d of the long-range protein-protein attractive potential, which are free fitting parameters of each of the SAXS investigated curves.

These temperature trends firstly show that 2 g/L insulin molecules at pH=3, in the absence of modified-sugars (black curves), are mainly present in monomeric or dimeric state, with a minimum $x_1 \approx 0.8$ at $\approx 40°$ C and a maximum $x_1 \approx 1$ at the highest temperatures.

The relative densities of hydration water have been found very similar for each of the $N_s = 4$ IN states, with an average value of 1.06 ± 0.01. Unique fitted values of the average radius R_j of 1, 2, 4 and 6 state are (9.6±0.1) Å, (13.2±0.5) Å, (23.3±0.2) Å and (27.0±0.3) Å, respectively. The buffer contribution to the ionic strength is (4.1±0.2) mM.

Results of insulin without modified-sugar and the effects provided by each of the two modified-sugars are discussed in the next paragraphs.

3.2.4.9 Insulin without and with modified-sugar

Insulin in solution is mainly found in the form of monomer. The molar fraction of nominal IN monomers that remain in the monomeric state in solution is indeed $x_1 \approx 0.8$ (Fig. 3.29, panel **A**, black line) while the rest are forming dimers ($x_2 \approx 0.2$, Fig. 3.29, panel **B**, black line) and neither tetramers nor hexamers are found (Fig. 3.29, panels **C** and **D**, black lines). Both fractions x_1 and x_2 do not show a marked dependence on temperature, even if at around 60° C dimers disappears (x_2 tend to zero).

The addition of modified-sugars, in particular of one of them (EC101), induces a completely different behavior, with the prevalence of tetramers and hexamers that are negligible in the absence of modified-sugars. Because hexamers represent the best oligomers to store and stabilize the functional monomers, this findings suggest that EC101 can be a successful compound for storing insulin.

3.2.4.10 Insulin with EC312

Results show that the increase of the concentration of EC312 (red curves) determines a decrease of IN monomers in favor of dimers (Fig. 3.29, panels **A** and **B**, red lines). The major difference is visible at 0.1 M and 0.25 M, when the fraction of dimers, x_2, increases from ≈ 0.3 to ≈ 0.6 with a slight dependence on the temperature until 45° C, after which monomers slowly increase up to $x_1 \approx 0.8$. Tetramers and hexamers are not present in solution during the EC312 addition. This effect can be also observed in panels **I-K**: insulin in absence of modified-sugars shows the lowest value of $\bar{K}_{12}$, indicating that the protein tends to stay in the monomeric state, whereas the addition of EC312 yields to higher $\bar{K}_{12}$ values and lower $\bar{K}_{24}$ and $\bar{K}_{46}$ values, confirming that EC312 favors the propensity of insulin to be found as a dimer in solution. The stabilization of the dimer in the presence of EC312 is clear considering the values of the exchange constants $K_{\mathrm{ex}j}$ reported in panels **Q-T**: the monomer, the tetramer and the hexamer shows $K_{\mathrm{ex}j} > 1$, whereas for dimers $K_{\mathrm{ex}j} < 1$, suggesting a preferential solvation of the dimer with EC312 in respect to water. We underline that, despite this preference only slightly modifies the water fraction in the first hydration shell of the dimers (ϕ_2 has a minimum value of ≈ 0.992, panel **F**), this small effect is sufficient to provoke an important increase of the monomer-dimer effective equilibrium constant $\bar{K}_{12}$ (panel **I**). The trends of the IN-IN structure factors $S(q)$, as well as the ones of the corresponding pair potentials $u(r)$, reported in Figs. 3.38 and 3.39 of the Supplementary Materials (Section 3.2.6) (red lines), clearly show a prevalence of long-range attractive forces in respect to repulsive forces. We also note that, up to 0.1 M EC312, the trends are quite similar to the ones observed for IN in the absence of modified-sugar (black lines), without significant variations with T. Conversely, at 0.25 M EC312, the attractive interactions raise and become much more marked as the temperature increases.

3.2.4.11 Insulin with EC101

EC101 behaves in a totally different way from EC312. Although the lowest concentration of EC101 retains a small fraction of monomers in solution ($x_1 \approx 0.2$, Fig. 3.29 panel **A**, green curves), which does not change considerably as a function of T, when insulin is mixed with EC101, insulin is mainly present as a tetramer or a hexamer. At 0.25 M EC101 only tetramers are in solution ($x_4 \approx 1$), whereas at lower concentration, the EC101 causes the formation of hexamers ($x_6 \approx 0.5-0.6$ at 40° C, panel **D**), with the remaining percentage occupied mainly by tetramers and in small part by monomers. The increase of temperature determines a negative slope of the x_6 *vs.* T curve, leading to a decrease of hexamers in favor of tetramers. The equilibrium constants $\bar{K}_{12}$, $\bar{K}_{24}$, and $\bar{K}_{46}$ (panels **I**-**K**, green curves) are bigger than the ones of IN in the absence of EC101 (black curves), and they grow additionally at increasing concentration of EC101. To note, the higher value of $\bar{K}_{24}$ (panel **J**, green curves), which, according to describes how the equilibrium from dimers to tetramers changes in presence of EC101, confirms the prevalence of tetramer at 0.25 M EC101, as indicated in panel **C** (green curves). The exchange constants $K_{\mathrm{ex}j}$, reported in panels **Q**-**T** are found to be greater than 1 both for monomers and dimers and smaller than 1 for tetramers and hexamers. These results clearly show a preferential solvation of tetramers and hexamers with EC101 with respect to water and an opposite preference of monomers and dimers for water. This is the mechanism that shows the capability of EC101 in stabilizing tetramers and hexamers. Comparing the results obtained with different proteins, we could infer that the stabilizing effect of the tested compounds works as a specific binomial modified-sugar-protein. Although some sugars, like trehalose, are commonly known to be stabilizers for biological macromolecules, their effect is always related to the specific protein. In particular, when dealing with proteins that present oligomeric equilibria, compounds efficiency in stabilizing each particular species is to be tested, since it depends on many features characterizing the macromolecule (charge, cavities, exposed groups, flexibility, etc.). Concerning the protein-protein structure factors and the related pair potentials of IN in the presence of EC101 (Figs. 3.38 and 3.39 of the Supplementary Materials (Section 3.2.6), green lines), results show that the prevalence of long-range attractive forces at any EC101 concentration, which grow with temperature.

3.2.5 Modified sugars discussion and conclusions

We have shown that, by using an approach that includes both structural and thermodynamic features of a protein in solution, it is possible to extract from a batch of SAXS curves recorded at several conditions of temperature and protein as well as cosolvent concentrations crucial information regarding the stabilizing effects of cosolvents. The model we have developed focuses on the preferential water solvation properties over the surface of each of the distinct states that proteins can form in solution and shows how the modifications of these properties, due to the presence of a cosolvent, can provide changes on the distribution of protein molecules among the different states. Although SAXS experiments can only concern a limited number of conditions in terms of temperature and proteins or cosolvent concentration, most of the fitting parameters of our model do not refer to a specific experiment but to the whole set of thermodynamic laws that regulate the behavior of the protein system at any physical-chemical condition. An important consequence of this approach is the possibility to calculate the phase-diagram of the protein as a continuous function of temperature and cosolute concentration.

Phase-diagrams derived by the two sets of SAXS data that we have analyzed in this work are shown in Fig. 3.30, for MB in the presence of five ExtremoChem modified-sugar, and in Fig. 3.31 for IN in contact with two of these modified-sugars. These diagrams contain the same information provided by the plots of x_j shown in Figs. 3.27 (panels **A**-**C**) and 3.29 (panels **A**-**D**) but allow a more immediate visualization of the achieved results. To note, the solid lines represent the thermodynamic condition in which at least one x_j is 0.5. Regarding the MB case, Fig. 3.30 (panels **A** and **F**) shows that EC312 is the best stabilizing modified-sugar, since, at 0.25 M, it preserves the monomeric N-state (blue area) up to ≈ 65° C. On the other hand, we see that 0.25 M EC101 (panels **B** and **G**) stabilize the N-state as well as the folded and active dimeric I-state (gold area) against the unfolded U-state (magenta area). EC311 (panels **C** and **H**) looks similar to EC312, but at 10 g/L MB it better stabilizes the I-state. We also see that EC202 (panels **D** and **I**) determines the largest stabilization area of the I-state against the U-state. Finally, EC212 (panels **E** and **J**) results are similar to the EC312 ones, but with a more marked stabilization of the I-state at 10 g/L MB. More in general, it is worth to notice that the phase-diagrams of uncharged compounds (EC312, EC311 and EC212) are qualitatively similar and differ from the phase-diagrams of the two charged compounds (EC101 and EC202), which have an evident stabilization effect of the active I-state.

Regarding insulin, phase-diagrams shown in Fig. 3.31 confirm a totally different behavior in the presence of EC312 (panel **A**) with respect to EC101 (panel **B**): the former mainly stabilizes the monomer state, at least up to ≈ 0.2 M, the latter, at 0.25 M, promotes the tetramers, whereas at concentrations between 0.05 and 0.15 M and temperatures comprised between 25° and 50° C favors the presence of hexamers.

SAXS data contain also information regarding the long-range interactions of proteins, which our model is able to dissect. In the case of both MB at pH=5 and IN at pH=3 our result indicate the attractive forces dominate with respect to Coulumbian repulsion, in particular at the highest concentrations of modified-sugar and temperature. Although our SAXS q-range does not allow to clearly identify the presence and the structure of high molecular weight species, our data suggest that they would be present, probably as unspecific aggregates. Further experimental evidences will be necessary to confirm this aspect.

The overall results achieved with the present study suggest that synchrotron-based SAXS technique, combined with advanced data analysis methods, is an invaluable tool for obtaining a detailed picture of thermal stability, oligomer distribution and long-range interactions of proteins in the presence of cosolvents.

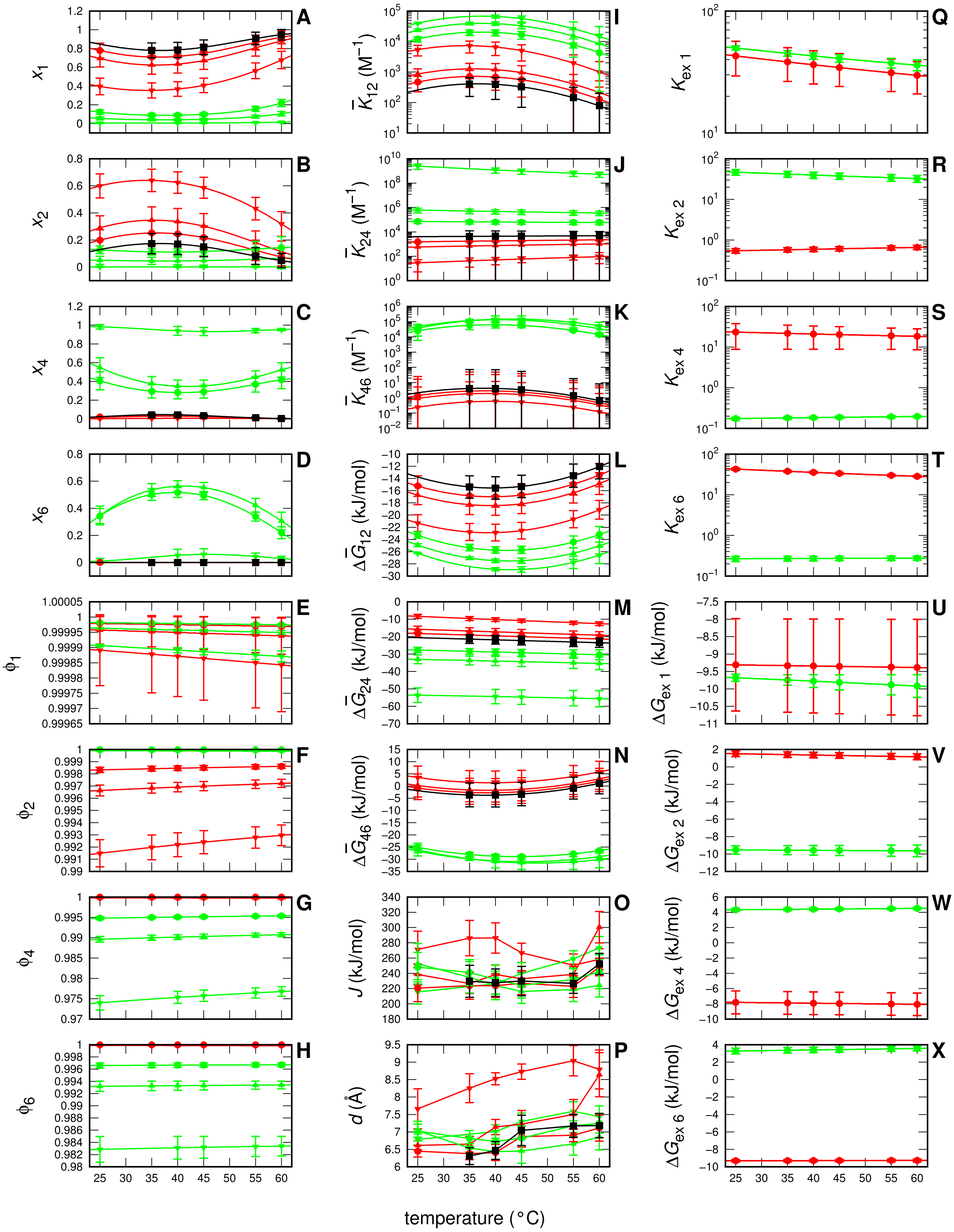


FIGURE 3.29: Temperature behaviors of the most relevant physical-chemical parameters obtained by the global-fit of 2 g/L IN SAXS curves shown in Fig. 3.28. Color refers to: no-modified-sugar (black), EC312 (red), EC101 (green). Thickness refers to: 0.05 M (thin), 0.10 M (intermediate), 0.25 M (thick). Point-type refers to: no-modified-sugar (square), 0.05 M (circle), 0.10 M (up-sided triangle), 0.25 M (down-sided triangle).

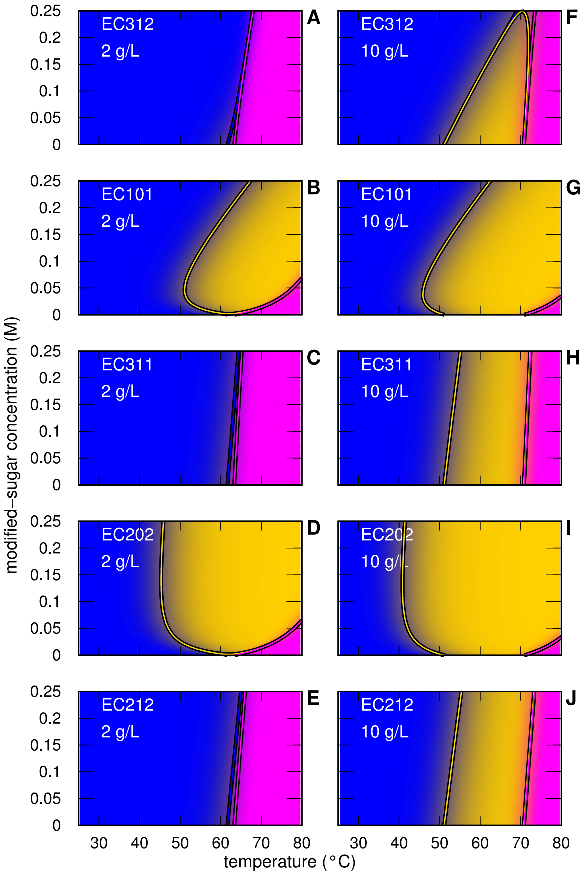


Figure 3.30: Temperature-modified-sugar-concentration phase-diagrams for MB in solution as obtained by the global-fit analysis of the SAXS curves. Panels in the same row refer to the same modified-sugar, as indicated, whereas left and right column refer to 2 and 10 g/L MB concentration. The color code of each condition has been calculated by mixing, according to the protein j-state distribution (x_j), the following pure colors assigned to each j-state: N (blue), I (gold) and U (magenta). Solid lines are the contour levels corresponding to $x_j = 0.5$ and their color has been assigned on the basis of the j-state.

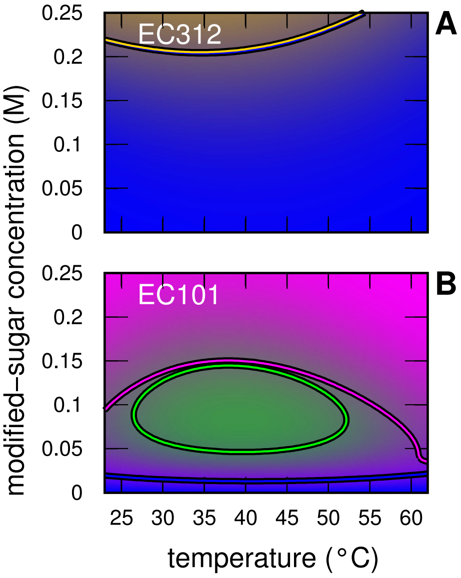


FIGURE 3.31: Temperature-modified-sugar-concentration phase-diagrams for 2 g/L IN in solution as obtained by the global-fit analysis of the SAXS. Panels **A** and **B** refer to EC312 and EC101, as indicated. The color code of each condition has been calculated by mixing, according to the protein j-state distribution (x_j), the following pure colors assigned to each j-state: monomers ($j = 1$, blue), dimers ($j = 2$, gold), tetramers ($j = 4$, magenta) and hexamers ($j = 6$, green). Solid lines are the contour levels corresponding to $x_j = 0.5$ and their color has been assigned on the basis of the j-state.

3.2.6 Supplementary Materials

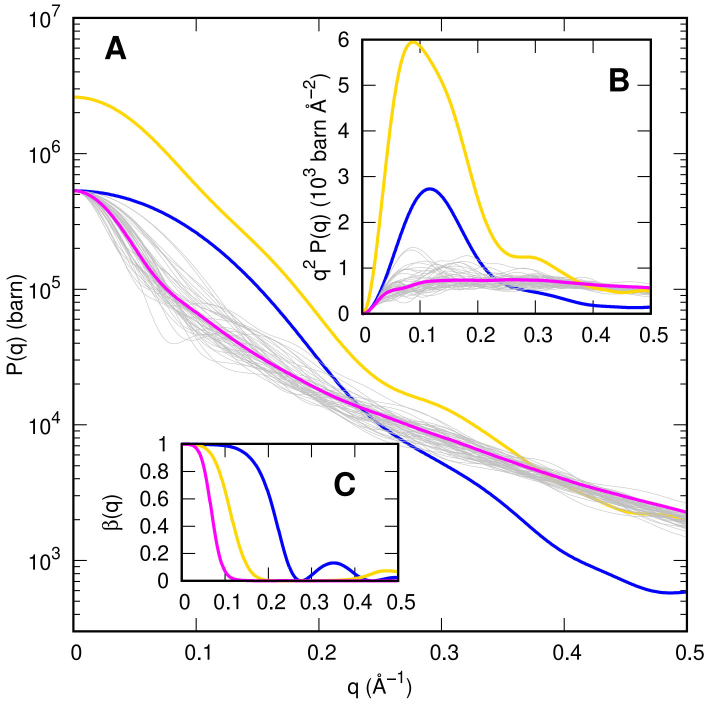


FIGURE 3.32: Form factors of the $N_s = 3$ states of MB calculated with the SASMOL method [58]. The blue curve represents the form factor of the native state (N) obtained on the basis of the PDB entry 1wla [44]. The gold curve is the form factor of the intermediate (I) dimeric state calculated from the PDB entry 3vm9 [48]. Grey curves are the form factors of the 50 conformations obtained by FOX [45], whose average, corresponding to the the unfolded state (U), is represented by the magenta curve. Form factors are shown in the semi-logarithmic (panel **A**) and in the Kratky (panel **B**) plots. Coupling functions are reported in panel **C**.

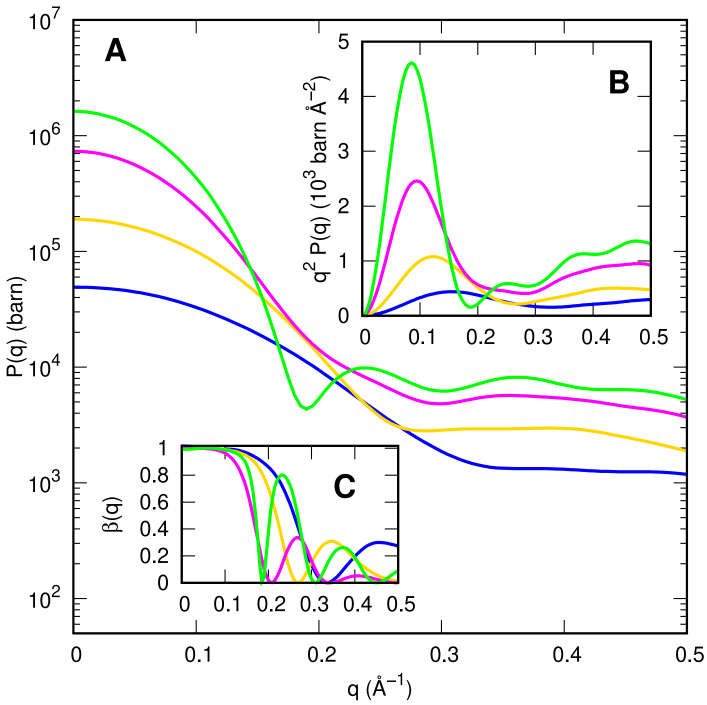


FIGURE 3.33: Form factors of the $N_s = 4$ states of IN calculated with the SASMOL method. Curves of monomer (1), dimer (2), tetramer (4) and hexamer (6) states, calculated from the PDB entry 3aiy [51], by selecting chains A-B, A-D, A-H and A-L, respectivley, are shown in blue, gold, magenta and green, respectivley. Form factors are shown in the semi-logarithmic (panel **A**) and in the Kratky (panel **B**) plots. Coupling functions are reported in panel **C**.

buffer pH	G	C_{G} M	pH	MB Z_N	IN Z_1
5.00	–	–	5.00	17.6	
5.00	EC101	0.05	5.15	16.8	
5.00	EC101	0.10	5.25	16.4	
5.00	EC101	0.25	5.35	15.9	
5.00	EC202	0.05	5.10	17.0	
5.00	EC202	0.10	5.20	16.6	
5.00	EC202	0.25	5.30	16.1	
3.00	–	–	3.00		4.7
3.00	EC101	0.05	3.80		2.8
3.00	EC101	0.10	4.15		1.9
3.00	EC101	0.25	4.50		1.0

TABLE 3.8: Experimental pH values determined as a function of the concentration of EC101 or EC202 modified-sugar dissolved in 10 mM phosphate buffer at pH = 3 or pH = 5 and corresponding number of elementary charges for the N-state of myoglobin and for the 1-state of insulin.

Symbol	Description	Unit	Min	Max	Ref.
α_{w}	water thermal expansivity	10^{-4} K^{-1}	**2.5**		[38]
β_{w}	first derivative of water thermal expansivity	10^{-6} K^{-2}	**9.8**		[38]
α_{P}	protein thermal expansivity	10^{-4} K^{-1}	**1.15**		[25]
α_{G}	modified-sugar thermal expansivity	10^{-4} K^{-1}	**3.9**		[71]
$\nu^{\circ}_{\mathrm{W_b}}$	bulk water molar volume at $T_{\circ}$	L	**0.018**		[38]
ν°_{P}	MB monomer molar volume at $T_{\circ}$	L	**12.95**		[58][a]
$\nu^{\circ}_{\mathrm{G_b}}$	bulk modified-sugar molar volume at $T_{\circ}$	L	**0.12**		
I_0	ionic strength due to charged buffer molecules	mM	1	20	
$\Delta G^{\circ}_{\mathrm{W,nel},j_1 j_2}$	non-electrostatic reference Gibbs free energy change at the $j_1 j_2$ transition	$\mathrm{kJ\,mol^{-1}}$	-200	200	
$\Delta S^{\circ}_{\mathrm{W}j_1 j_2}$	reference entropy change at the $j_1 j_2$ transition	$\mathrm{J\,mol^{-1}}$	-200	200	
$\Delta C_{p\mathrm{W}j_1 j_2}$	heat capacity at constant pressure change at the $j_1 j_2$ transition	$\mathrm{J\,mol^{-1}}$	-10000	10000	
$\Delta G^{\circ}_{\mathrm{ex}j}$	modified-sugar water exchange reference Gibbs free energy change over the j-state	$\mathrm{kJ\,mol^{-1}}$	-10	10	
$\Delta S^{\circ}_{\mathrm{ex}j}$	modified-sugar water exchange reference entropy change over the j-state	$\mathrm{J\,mol^{-1}}$	-10	10	
$\Delta C_{p_{\mathrm{ex}j}}$	modified-sugar water exchange heat capacity at constant pressure change over the j-state	$\mathrm{J\,mol^{-1}}$	-10	10	
d_j	relative mass density of the j-state hydration water		0.9	1.15	[87]
R_N	average radius of N-MB state	Å	12	18	
R_I	average radius of I-MB state	Å	20	40	
R_U	average radius of U-MB state	Å	30	70	
Z_N	number of elementary charges of N-MB		see Table 3.8		
J	protein-protein contact energy	$\mathrm{kJ\,mol^{-1}}$	0	1000	[83]
d	attractive potential scale length	Å	0.5	100	[83]

TABLE 3.9: Overview of the model parameters and their validity range used in the global-fit analysis of $N_c = 92$ SAXS curves of MB samples. Fixed parameters are shown in bold. [a] Value obtained with the SASMOL method. The total number of fitting parameters are 242, corresponding to ≈ 2.6 parameters per curve.

Symbol	Description	Unit	Min	Max	Ref.
α_{w}	water thermal expansivity	10^{-4} K^{-1}	**2.5**		[38]
β_{w}	first derivative of water thermal expansivity	10^{-6} K^{-2}	**9.8**		[38]
α_{P}	protein thermal expansivity	10^{-4} K^{-1}	**1.15**		[25]
α_{G}	modified-sugar thermal expansivity	10^{-4} K^{-1}	**3.9**		[71]
$\nu^{\circ}_{\mathrm{W_b}}$	bulk water molar volume at $T_{\circ}$	L	**0.018**		[38]
ν°_{P}	IN monomer molar volume at $T_{\circ}$	L	**4.32**		[58][a]
$\nu^{\circ}_{\mathrm{G_b}}$	bulk modified-sugar molar volume at $T_{\circ}$	L	**0.12**		
I_0	ionic strength due to charged buffer molecules	mM	1	20	
$\Delta\bar{G}^{\circ}_{\mathrm{W,nel},12}$	non-electrostatic reference Gibbs free energy change at the 12 transition (Eq. 26)	$\mathrm{kJ\,mol^{-1}}$	-50	0	[36]
$\Delta\bar{G}^{\circ}_{\mathrm{W,nel},24}$	non-electrostatic reference Gibbs free energy change at the 24 transition (Eq. 26)	$\mathrm{kJ\,mol^{-1}}$	-30	30	[36]
$\Delta\bar{G}^{\circ}_{\mathrm{W,nel},46}$	non-electrostatic reference Gibbs free energy change at the 46 transition (Eq. 26)	$\mathrm{kJ\,mol^{-1}}$	-30	30	[36]
$\Delta\bar{S}^{\circ}_{\mathrm{W}j_1j_2}$	reference entropy change at the j_1j_2 transition (Eq. 26)	$\mathrm{J\,mol^{-1}}$	-40	40	
$\Delta\bar{C}_{p\mathrm{W}j_1j_2}$	heat capacity at constant pressure change at the j_1j_2 transition (Eq. 26)	$\mathrm{J\,mol^{-1}}$	-10000	10000	
$\Delta G^{\circ}_{\mathrm{ex}j}$	modified-sugar water exchange reference Gibbs free energy change over the j-state	$\mathrm{kJ\,mol^{-1}}$	-10	10	
$\Delta S^{\circ}_{\mathrm{ex}j}$	modified-sugar water exchange reference entropy change over the j-state	$\mathrm{J\,mol^{-1}}$	-10	10	
$\Delta C_{p_{\mathrm{ex}j}}$	modified-sugar water exchange heat capacity at constant pressure change over the j-state	$\mathrm{J\,mol^{-1}}$	-10	10	
d_j	relative mass density of the j-state hydration water		0.9	1.15	[87]
R_1	average radius of 1-IN state	Å	10	20	
R_2	average radius of 2-IN state	Å	10	30	
R_4	average radius of 4-IN state	Å	15	35	
R_6	average radius of 6-IN state	Å	15	45	
Z_1	number of elementary charges of 1-IN		see Table 3.8		
J	protein-protein contact energy	$\mathrm{kJ\,mol^{-1}}$	0	1000	[83]
d	attractive potential scale length	Å	0.5	100	[83]

TABLE 3.10: Overview of the model parameters and their validity range used in the global-fit analysis of $N_c = 40$ SAXS curves of IN samples. Fixed parameters are shown in bold. [a] Value obtained with the SASMOL method. The total number of fitting parameters are 122, corresponding to ≈ 3.0 parameters per curve.

3.2.6.1 Numerical determination of x_j and ϕ_j

By assuming small values of X_{P} and x_{G}, in a 0th approximation, on the basis of Eq. 12 and 17, we fix $X_{\mathrm{W_b}}^{(0)} \simeq (1 - X_{\mathrm{P}})(1 - x_{\mathrm{G}})$ and $x_{\mathrm{G_b}}^{(0)} \simeq x_{\mathrm{G}}(1 - X_{\mathrm{P}})$. We calculate the charge Z_j of any protein state and, by assuming in the 0th approximation $I_{\mathrm{ci}} \simeq 0$, we calculate the related electrostatic free energy $G_{\mathrm{W,el},j}$ from Eq. 14. Hence from Eq. 8 we derive a 0th approximation of ϕ_j according to $\phi_j^{(0)} \simeq K_{\mathrm{ex}j}(1 - x_{\mathrm{G_b}}^{(0)})/(K_{\mathrm{ex}j} - (K_{\mathrm{ex}j} - 1)x_{\mathrm{G_b}}^{(0)})$. Subsequently, by entering $\phi_j^{(0)}$ in Eqs. 15-16, we derive $<\nu>^{(0)}$ and $C_{\mathrm{P}}^{(0)}$. By substituting these values into Eq. 10, we derive a 0th approximation of $K_{1j}^{(0)}$ and of the fraction of j-proteins $x_j^{(0)} \simeq \alpha_j (K_{1j}^{(0)} x_1^{(0)})^{\alpha_j} (C_{\mathrm{P}}^{(0)})^{\alpha_j - 1}$ (with $j = 2, N_s$), all depending on the unknown variable $x_1^{(0)}$. By entering the values of $x_j^{(0)}$ in Eq. 3, we obtain an equation with a unique variable $x_1^{(0)}$, comprised between 0 and 1, which can be numerically solved. Subsequently, by entering the values of $x_j^{(0)}$ in Eqs. 12-16, we derive a 1st approximation of the variables $X_{\mathrm{W_b}}^{(1)}$, $x_{\mathrm{G_b}}^{(1)}$, $<\nu>^{(1)}$, $C_{\mathrm{P}}^{(1)}$ and I_{ci}. Hence, again from Eq. 8, we derive 1st approximations of $\phi_j^{(1)}$ and, from these ones, 1st approximations of $K_{1j}^{(1)}$ and, finally of $x_j^{(1)}$. The iteration is repeated until $|x_j^{(i)} - x_j^{(i-1)}| < \tau$ and $|\phi_j^{(i)} - \phi_j^{(i-1)}| < \tau$, where τ is a small value in the order of 10^{-10}.

3.2.6.2 Minimization of the merit function

The merit function $\mathcal{H}$ is minimized by using two numerical methods, both integrated in the GENFIT software. The first method exploits a Simulated Annealing (SA) procedure. All the fitting parameters are randomly moved within their validity range: a move that decreases $\mathcal{H}$ ($\Delta\mathcal{H} \leq 0$) is always accepted, whereas a move that increases $\mathcal{H}$ is accepted with probability $e^{-\Delta\mathcal{H}/T^*}$, where T^* is a temperature-like parameter. The initial value of T^* is set to $\mathcal{H}_0/2$, where $\mathcal{H}_0$ is the value of the merit function corresponding to a first random choice of the fitting parameters. This high value of T^* ensures that the minimization starts by accepting most of the random moves, avoiding to get stuck in local minima. Every N_c random moves of the fitting parameters (referred to as a sub-run), T^* is decreased by a constant factor (i.e. the system is "cooled") so that the path along the space of the parameters is guided toward the global minimum of $\mathcal{H}$. We have chosen a total of $N_s = 50$ sub-runs formed by $N_c = 50$ random moves. The second method is based on the well-known simplex (SX) algorithm for solving numerically linear programming problems [47] and is applied to the set of fitting parameters that has been obtained by the SA method. The coupled SA-SX procedure is repeated for $N_I = 20$ independent iterations and, for each repetition, all the SAXS curves are randomly sampled from a Gaussian centered around the experimental value $\frac{d\Sigma}{d\Omega}_{k,\mathrm{expt}}(q_i)$ with standard deviation $\sigma_k(q_i)$ [81]. In this manner, a robust determination of the mean value $< X >$ and the standard deviation σ_X of each fitting parameters is achieved.

3.2.6.3 Calculation of $S(q)$

The modification of the PY structure factor $S_0(q)$ caused by the two Yukawian terms of the pair potential within the RPA approximation is

$$S(q) = \frac{S_0(q)}{1+\beta <n> S_0(q)[U_{\mathrm{YC}}(q)+U_{\mathrm{YA}}(q)]}, \tag{3.32}$$

$$S_0(q) = \left\{1 - \frac{12\eta[\eta(3-\eta^2)-2]}{(1-\eta)^4}\frac{j_1(2qR)}{2qR}\right\}^{-1}. \tag{3.33}$$

These equations contain the protein volume fraction, $\eta = <n> \frac{4}{3}\pi R^3$, and the isotropic Fourier Transform of the Yukawa potential, which reads $U_{\mathrm{Y}k}(q) = 4\pi B_{1k}(B_{2k}\sin(2qR) + q\cos(2qR))/(q^3 + qB_{2k}^{\,2}))$, with $k = \mathrm{C}$ and A and $\beta = 1/(k_B T)$. The average number density $<n>$ is calculated within the thermodynamic model according to $<n> = <\alpha^{-1}> n$, with $n = C_{\mathrm{P}}/N_A$. The function $j_1(x)$ is the 1^{st} order spherical Bessel function.

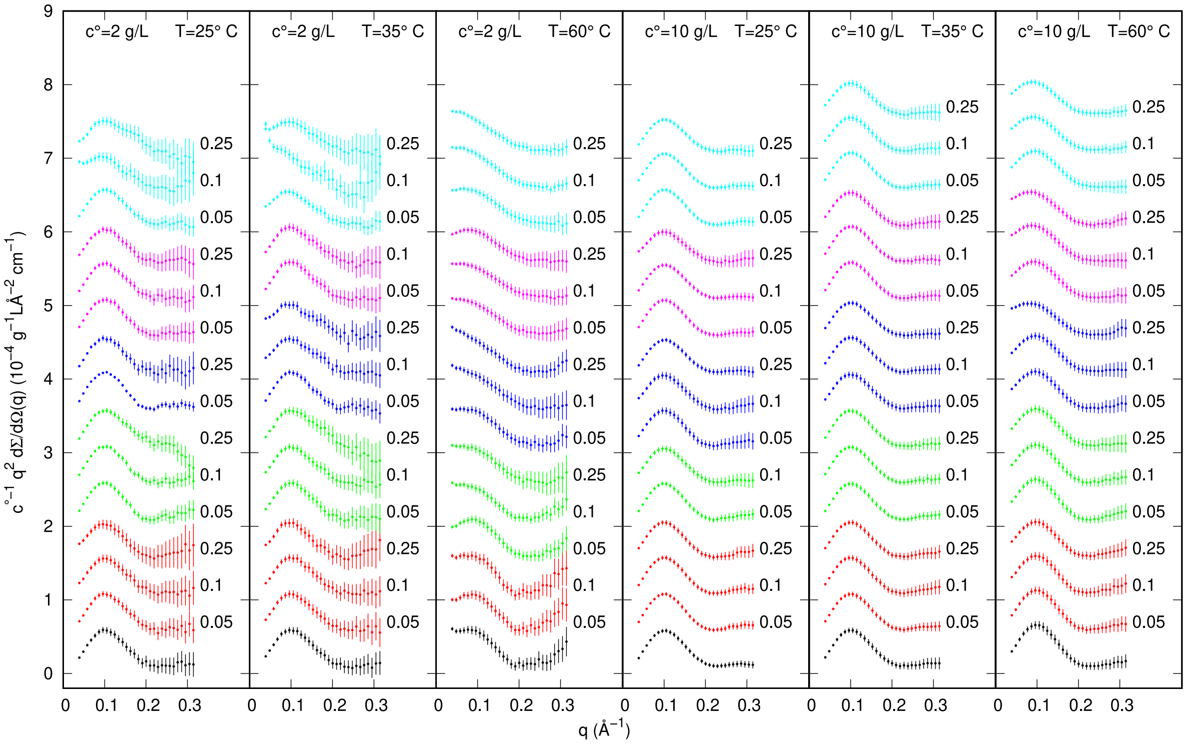


FIGURE 3.34: Kratky plots of the experimental SAXS curves of MB in 10 mM phosphate buffer (pH = 5) with and without ExtremoChem modified-sugar. Colors refer to the following conditions: no-modified-sugar (black), EC312 (red), EC101 (green), EC311 (blue), EC202 (magenta), EC212 (cyan). Whenever present, the modified-sugar concentration is reported on the right side of each curve in molar unit. Each column refers to a fixed temperature and MB concentration, as indicated on the top. Curves have been divided by the nominal w/v MB concentration ($c°$) and, in the same column, have been scaled by the factor 1 (in the unity of the y-axis). Experimental standard deviations are reported as error bars every 10 points, for clarity.

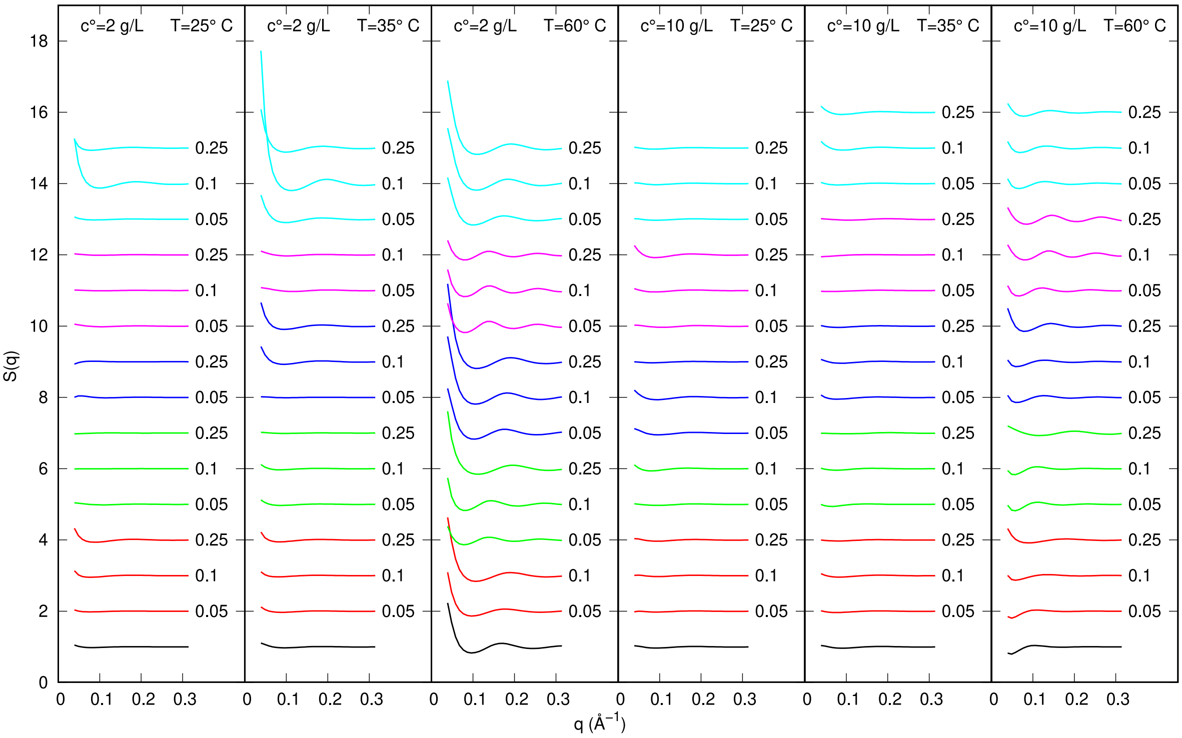


FIGURE 3.35: Protein-protein structure factors obtained by the analysis of SAXS data of MB samples shown in Figure 3. Curves are scaled by a factor 1 for clarity.

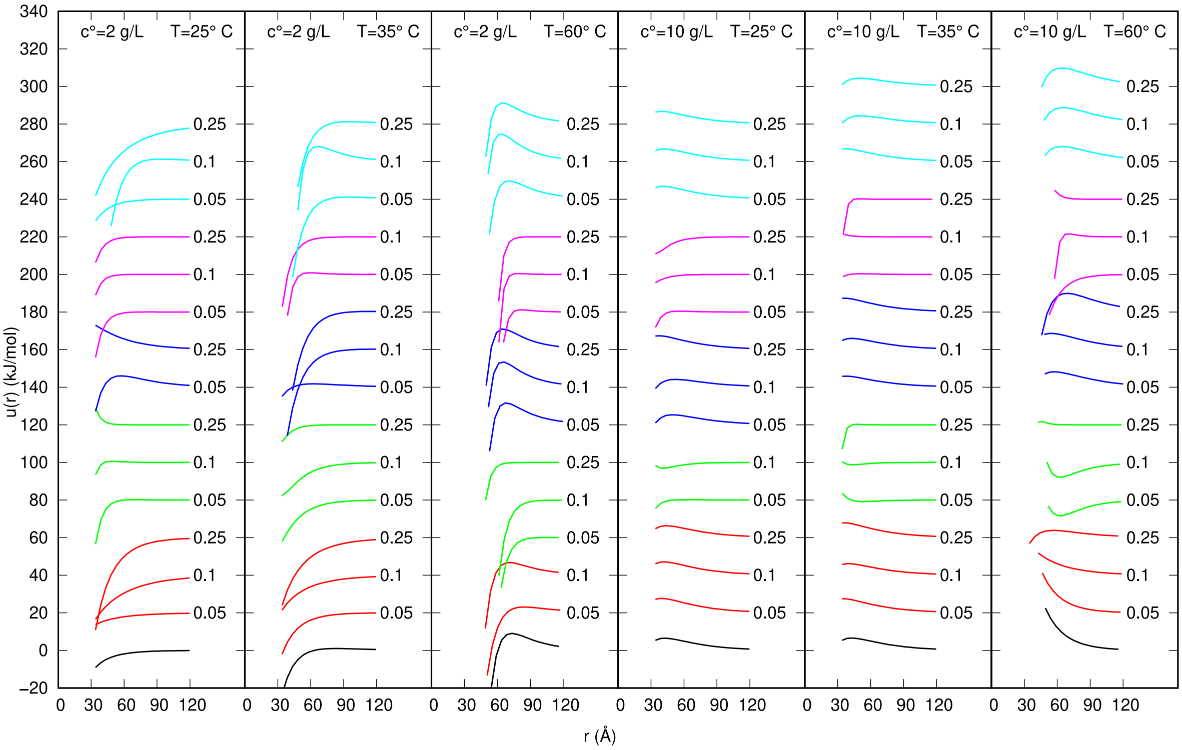


FIGURE 3.36: Protein-protein pair potentials obtained by the analysis of SAXS data of MB samples shown in Figure 3. Curve are scaled by a factor 20 kJ/mol for clarity.

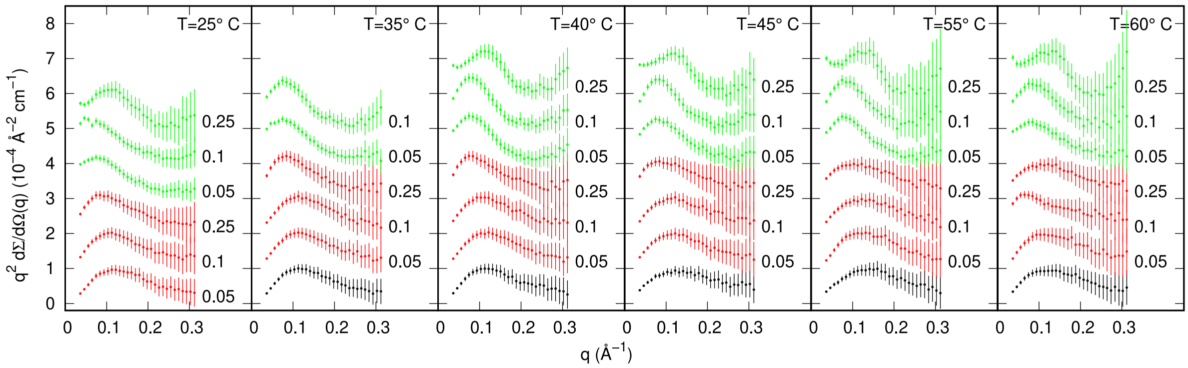


FIGURE 3.37: Kratky plots of the experimental SAXS curves of 2 g/L IN in 10 mM phosphate buffer (pH = 3) with and without ExtremoChem modified-sugar. Colors refer to the following conditions: no-modified-sugar (black), EC312 (red), EC101 (green). Whenever present, the modified-sugar concentration is reported on the right side of each curve in molar unit. Each column refers to a fixed temperature, as indicated on the top. Curves in the same column have been scaled by the factor 1 (in the unity of the y-axis). Experimental standard deviations are reported as error bars every 10 points, for clarity.

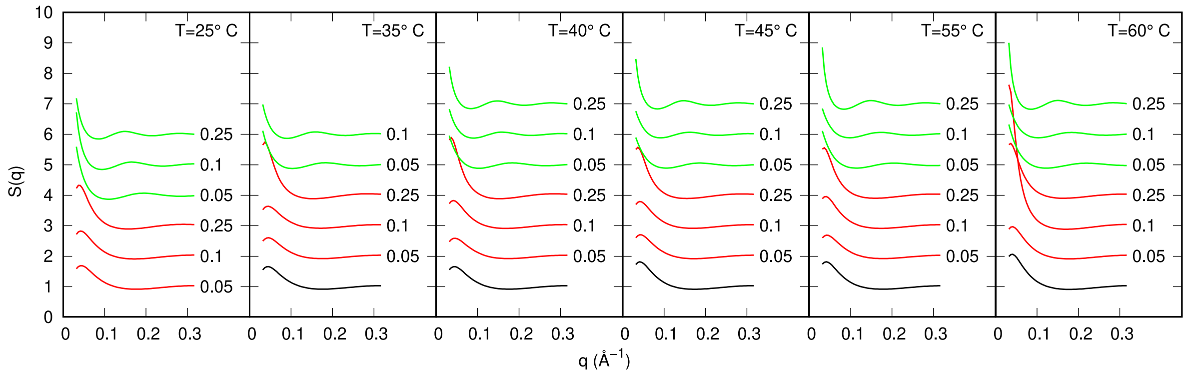


FIGURE 3.38: Protein-protein structure factors obtained by the analysis of SAXS data of IN samples shown in Figure 5. Curves are scaled by a factor 1 for clarity.

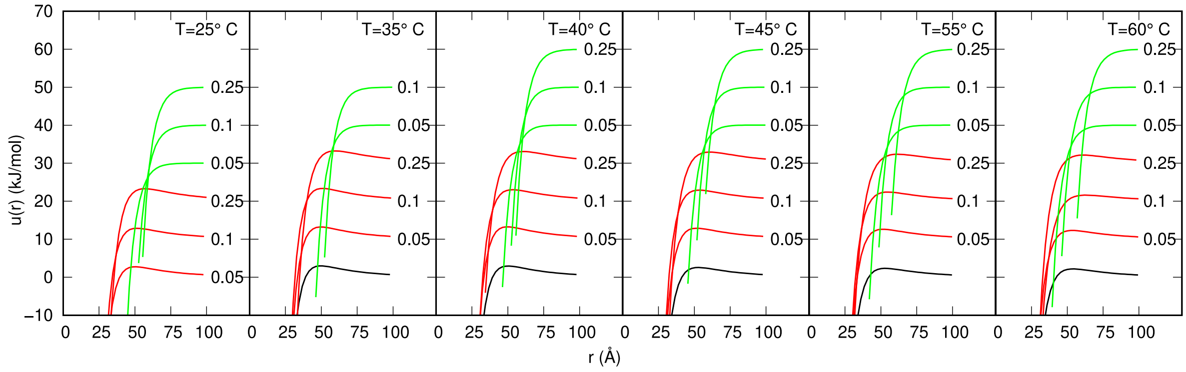


FIGURE 3.39: Protein-protein pair potentials obtained by the analysis of SAXS data of IN samples shown in Figure 5. Curve are scaled by a factor 10 kJ/mol for clarity.

3.3 SAXS/WAXS investigations of microfibers degradation processes under simulated solar irradiation

3.3.1 Fibers production and preparation

We have investigated five kinds of fibers, four of them were textile fibers (cotton, linen, polyester, polyamide) and were supplied by IPCB-CNR (Pozzuoli, Napoli, Italy), whereas the cellulose acetate fibers were prepared at the DISVA laboratory of the Polytechnic University of Marche (Ancona, Italy), by using a cryostat to cut non-smoked cigarette butts with 5 μm thickness. For each type of fiber, one aliquote was dissolved in 100 mL of artificial sea water (ASW, Instant Ocean®Sea Salt at 35 psu, 1 psu being 1 g per kg of water) and a second aliquote was dissolved in 100 mL of fresh water (tap water), both filtered at 0.2 μm pore size filters. For all the 10 samples, the w/v fiber concentration was 1 g/L. The bottles filled with fibers had the cap made of quartz because (Fig. 3.41) it does not absorb the UV component of the solar radiation (200-400 nm, Fig. 3.40), which is the component mostly responsible of material degradation processes, allowing it to enter the container and reach the fibers. Besides, each plastic part of the bottles was accurately covered by aluminum to reflect the simulated solar radiation elsewhere and avoid the risk of plastic degradation inside the samples. In order to simulate the degradation processes that take place in waterly environments, the bottles were additionally put over a rotating plate to take into account the wave motion and two ventilators were used to mitigate the heat flow generated by all the other component of the solar simulator (Fig. 3.42). Other two bottles, on filled with 100 mL of ASW and another with 100 mL of fresh water, were left free of fibers and used as controls. To simulate one year of solar irradiation, the average yearly solar radiance, 5% of which is due to UV radiation, was considered. The number of solar simulator exposure days, corresponding to the average yearly solar radiance, calculated by taking into account the solar simulator power (Table 3.11), was ≈ 42.

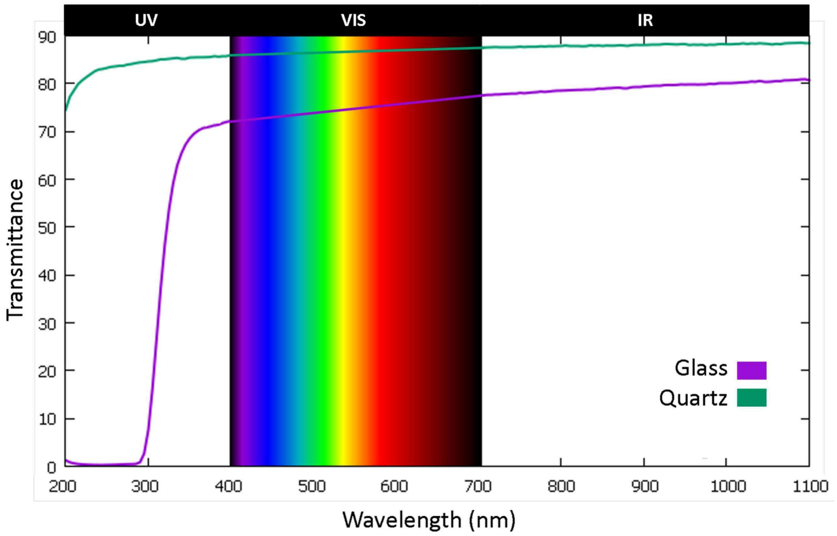


FIGURE 3.40: Comparison between glass and quartz transmittance considering the spectrum of the solar radiation.

Mean solar irradiance (kWh/m^2)	1200
5% UV component (kWh/m^2)	60
UV component of solar simulator power (kW/m^2)	0.06
Exposition time per day (h)	24
Total energy per day (kWh/m^2)	1.44
Total exposition time to simulate 1 year of solar exposure (days)	41.67

TABLE 3.11: Physical parameters considered for the evaluation of the fibers exposure time under the solar simulator apparatus.

FIGURE 3.41: Modified caps of the bottles with the top surface made of quartz (left) and bottles ready to be filled with fibers (center and right).

FIGURE 3.42: Solar simulator used during the experiment, kindly offered by the Department of Industrial Engineering and Mathematical Sciences at the Polytechnic University of Marche (Ancona, Italy).

3.3.2 Fibers sampling

Samplings were done regularly during the 42 days of solar simulator exposure. In particular, micropipettes with the tips half cut were used, in order to have a bigger diameter hole and suck a more homogeneous quantity of sample. Some pictures were taken during the experiment to attest any possible quality change in the samples (Fig. 3.43).

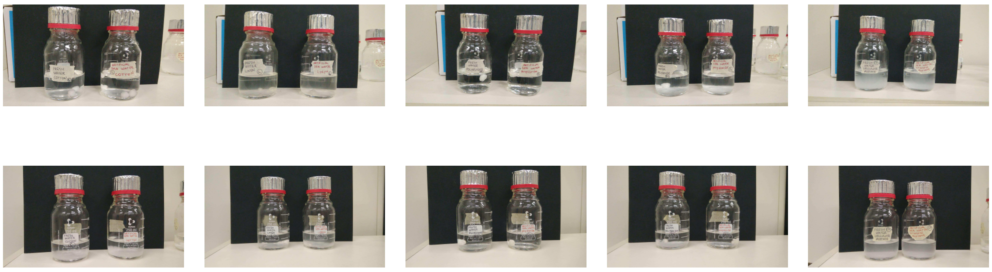


FIGURE 3.43: The first row shows samples before the irradiation process, whereas the second row show the same bottles after 42 days of solar simulation.

All the fibers, except for cellulose acetate, when put inside the water bottles tended to aggregate and form little balls of microfibers randomly arranged, making a homogeneous irradiation difficult to be achieved. For this reason, we decided to intervene and detangle all the floating yarns using a laboratory tweezer accurately washed and dried with compressed air. However, after this operation not all the microfibers remained separated and once the rotating plate was switched on they started the aggregation process again. Just cotton, among all the other fibers, kept its disaggregation state. Another visible change can be found by looking at the linen bottle where microfibers gradually changed their colour from yellowish to white.

3.3.3 SAXS/WAXS experiments

SAXS and WAXS experiments were performed at the Austrian SAXS beamline of Elettra synchrotron in Trieste in December 2021. Measurements were carried out at room temperature using the SAXS cell reported in Fig. 3.44. The q modulus of the scattering vector was fixed between 0.01 and 0.4 Å^{-1}, for SAXS, and between 1 and 2 Å^{-1} for

WAXS. For each sample eighteen frames were collected, treated with FIT2D, and then the data reduction was applied to obtain the final scattering curves that are reported in the sections below.

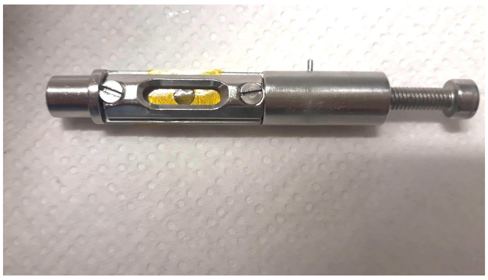


Figure 3.44: SAXS cell used during the SAXS/WAXS experiment at Elettra in December 2021. The solid sample of the fibers was adjusted inside the cell and then irradiated by X-ray.

3.3.4 Microfibers model

Microfibers are composed of several fibers of decreasing dimensions twisted one to another. One single microfiber is composed of other fibers in the nanometer range (see Fig. 3.45). According to literature [52], in this thesis we call the former *macrofibrils* and the latter *microfibrils*. Both of them have dimensions that fall in typical nanometric investigation range of SAXS (1-100 nm). In fact, to investigate the whole microfiber with X-ray scattering, it is necessary to use the ultra small-angle X-ray scattering technique (USAXS). As a consequence, the SAXS and WAXS models developed in this thesis only deal with the internal structure of the microfibers under investigation. In particular, to better understand the models that we used, we can consider the cross section of one single macrofibril as composed of several microfibrils enclosed inside of it, and surrounded by water (Fig. 3.46).

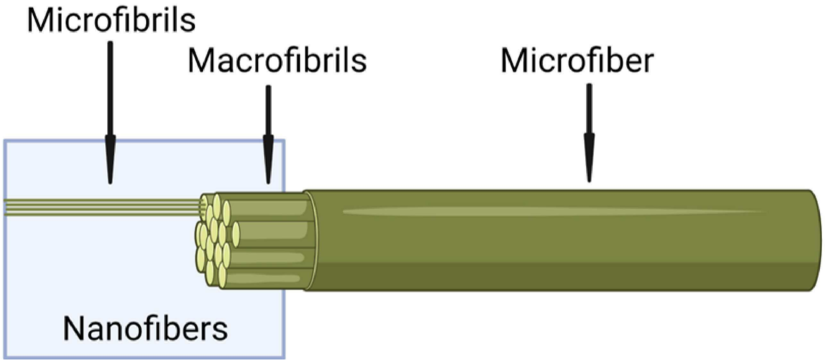


FIGURE 3.45: Representation of the internal and external structure of one microfiber.

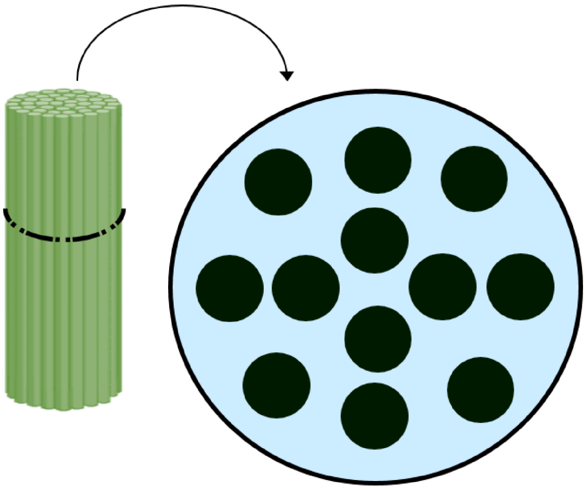


FIGURE 3.46: Internal representation of one macrofibril with several microfibrils enclosed inside it.

3.3.5 SAXS and WAXS models

3.3.5.1 SAXS model

SAXS curves of microfibers have been fitted with the worm-like model with excluded volume effects, developed by Pedersen and Schurtenberger [60], combined with the cross section of locally parallel cylinders, representing the microfibrils, organized in a two-dimensional finite and polydispersed hexagonal lattice, with paracrystalline lattice distortion of the second kind [43, 26, 85]. This model, inspired by the method of Penttilä et al. [61], has been implemented in the software GENFIT [81]. In detail, the macroscopic differential scattering cross section is written as

$$\frac{d\Sigma}{d\Omega}(q) \quad = \quad \kappa P_{\mathrm{wl}}(q) P_{\mathrm{cs}}(q) + B \tag{3.34}$$

where κ is a scaling factor and B is a flat background. As already discussed in Section 3.1.8, they are unimportant but necessary parameters in order to deal with SAXS data in arbitrary unit. $P_{\mathrm{wl}}(q)$ is the form factor of the infinitely thin worm-like chain: it depends on the statistical segment (Kuhn) length b, which can be considered the average length of the cylinder bundle (Fig. 3.48), and by the number of statistical segments n_b. Notice that the contour length L, which represents the length of the fully extented chain, is simply given by $L = n_b b$ (Fig. 3.47.

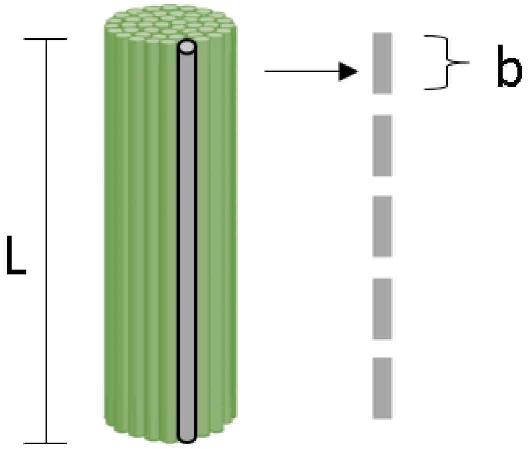


FIGURE 3.47: Macrofibril worm-like model representation.

The mean form factor of the macrofibril cross section, $P_{\mathrm{cs}}(q)$, is calculated as the integral, over the angle α around the axis of the cylinder bundle (Fig. 3.48), of the cross section $P_{\mathrm{cs}}(\mathbf{q}_{\|})$ that depends on the two-dimensional scattering vector in the plane perpendicular to the axis of the macrofibril, $\mathbf{q}_{\parallel} = (q\cos\alpha, q\sin\alpha)$, according to

$$P_{\rm cs}(q) \;=\; \frac{1}{2\pi}\int_0^{2\pi} P_{\rm cs}(\mathbf{q}_{\parallel})d\alpha. \tag{3.35}$$

In turn, for an array of cylinders with poly-dispersed radius, the function $P_{\rm cs}(\mathbf{q}_{\parallel})$, according to the finite paracrystal theory, is

$$P_{\rm cs}(\mathbf{q}_{\parallel}) \;=\; <A^2(q)>_R + <A(q)>_R^2 \left[S_{\rm cc}(\mathbf{q}_{\parallel}) - 1\right] \tag{3.36}$$

In this equation, $A(q)$ is the SAXS amplitude of an infinitely long cylinder with radius R and homogeneous electron density [28, 85],

$$A(q) \;=\; 2\pi R^2 \frac{J_1(qR)}{qR} \tag{3.37}$$

$J_1(x)$ being the Bessel function of the first kind. The two averages over the radius R shown in Eq. 3.36 are calculated assuming a Gaussian distribution function, centered in R_0 and with standard deviation $\sigma_R = g_R R_0$,

$$<A(q)>_R \;=\; \int_0^{\infty} A(q)\, p(R) dR \tag{3.38}$$

$$<A^2(q)>_R \;=\; \int_0^{\infty} A^2(q)\, p(R) dR \tag{3.39}$$

where the normalized Gaussian distribution is

$$p(R) = \frac{e^{-(R-R_0)^2/(2g_R^2R_0^2)}}{\int_0^{\infty} e^{-(R-R_0)^2/(2g_R^2R_0^2)} dR} \tag{3.40}$$

The function $p(R)$ allows to calculate the average radius,

$$< R >= \int_0^\infty R\, p(R) dR \tag{3.41}$$

Notice that, since the integral is calculated between 0 and ∞, $< R >$ could be different from R_0. In practice, all the integrals over R (Eqs. 3.38-3.41) are numerically calculated with the Simpson rule using 30 points in the range comprised between $\min\{0, R_0(1-3g_R)\}$ and $R_0(1+3g_R)$. In Eq. 3.36, the function $S_{\text{cc}}(\mathbf{q}_\|)$ represents the finite and poly-dispersed paracrystal structure factor of a hexagonal lattice. This function is calculated by the following expression [43, 26, 85],

$$\begin{aligned} S_{\text{cc}}(\mathbf{q}_\|) &= \sum_{i=-2\sigma_p}^{2\sigma_p} p_i^2 \prod_{k=1}^{2} S_{\text{PT}}(F_k, N_a + i) & (3.42) \\ S_{\text{PT}}(F_k, N_a + i) &= \Re \left[\frac{1 + F_k}{1 - F_k} - \frac{2F_k(1 - F_k^{N_a+i})}{(N_a + i)(1 - F_k)^2} \right] & (3.43) \end{aligned}$$

where $F_k = e^{-\frac{1}{2}(a g_a q)^2} e^{\imath \mathbf{q}_\| \cdot \mathbf{a}_k}$. The hexagonal unit cell vectors are $\mathbf{a}_1 = a\,\hat{\mathbf{x}}$ and $\mathbf{a}_2 = a(\cos\frac{\pi}{3}\,\hat{\mathbf{x}} + \sin\frac{\pi}{3}\,\hat{\mathbf{y}})$ ($\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors along x and y directions). Relevant parameters of the model, apart the lattice length a, are the number of unit cells N_a along each of the two directions $\mathbf{a}_1$ and $\mathbf{a}_2$, and the so-called distortion parameter g_a, defined as $g_a = \sigma_a/a$, σ_a being the Gaussian standard deviation of the paracrystal isotropic distortion. The polydispersity over N_a should be introduced in order to eliminate "intrinsic" oscillations of the monodisperse paracrystalline structure factor at low q that have never seen in experimental data [26]. Sampling points are weighed by a discrete Gaussian distribution

$$p_i = \frac{e^{-\frac{1}{2}i^2/\sigma_p}}{\sum_{i=-2\sigma_p}^{2\sigma_p} e^{-\frac{1}{2}i^2/\sigma_p^2}} \tag{3.44}$$

with standard deviation

$$\sigma_p = \begin{cases} \frac{1}{2}(N_a - 1) & N_a < 5 \\ N_a^{1/2} & N_a \geq 5 \end{cases} \tag{3.45}$$

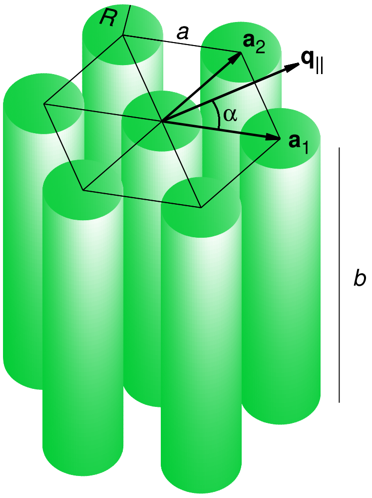


Figure 3.48: A representation of a bundle of parallel cylinders, with radius R and length b, packed in a hexagonal array with lattice parameter a and unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. The two-dimensional vector $\mathbf{q}_{\|}$ in the plane perpendicular to the bundle and forming an angle α with the x-axis is also represented.

3.3.5.2 WAXS model

The isotropic WAXS data that show diffraction peaks have been analyzed according, in general, to methods and results of Refs. [50], [18], [62] and [86], all within the finite three-dimensional paracrystalline theory [43, 26], which have been entered in the GENFIT software [81]. The macroscopic WAXS differential scattering cross section of the macrofibrils, obtained by subtracting the WAXS signal due to the the solvent (ASW or fresh water), is written as a combination of two paracrystalline systems; a Gaussian function and a power function are added, to take into account the contribution of the amorphous regions,

$$\frac{d\Sigma}{d\Omega}(q) \quad = \quad \kappa_1 \sum_{i=1}^{2} x_i S_{i,\mathrm{PT}}(q) + \kappa_2 e^{-(q-q_0)^2/(2\sigma_2^2)} + \kappa_3 q^{-\alpha_3} \tag{3.46}$$

In this equation κ_1, κ_2 and κ_3 are arbitrary scaling factors, x_1 and x_2 (with $x_1+x_2 = 1$) are the relative weights of the two paracrystalline systems, q_0 and σ_2 are the center and the standard deviation of the Gaussian function and α_3 is the exponent of the power function. The function $S_{i,\mathrm{PT}}(q)$ represents the orientational average of randomly distributed crystals having: unit cell lengths a_i, b_i and c_i; unit cell angles α_i, β_i and γ_i; number of cells along the unit vectors $N_{a,i}$, $N_{b,i}$ and $N_{c,i}$; paracrystal distortion parameters $g_{a,i}$, $g_{b,i}$ and $g_{c,i}$. On these premises, the unit cell vectors are: $\mathbf{a}_i = a_i(1,0,0)$; $\mathbf{b}_i = b_i(\cos\gamma_i, \sin\gamma_i, 0)$; $\mathbf{c}_i = c_i(\cos\beta_i, (\cos\alpha_i - \cos\beta_i\cos\gamma_i)/\sin\gamma_i, (1-\cos^2\beta_i - ((\cos\alpha_i - \cos\beta_i\cos\gamma_i)/\sin\gamma_i)^2)^{1/2}$. The random orientation of the paracrystal is calculated according to

$$S_{i,\mathrm{PT}}(q) = \frac{1}{N_u}\sum_{k=1}^{N_u} S_{i,\mathrm{PT}}(q\mathbf{u}_k) \tag{3.47}$$

where $\mathbf{u}_k$ is the k^{th} of N_u efficiently chosen points along the surface of a unit sphere, according to Ref. [93]. In practice, we have chosen the set of $N_u = 801$ points reported in Ref. [92] and represented in Fig. 3.49.

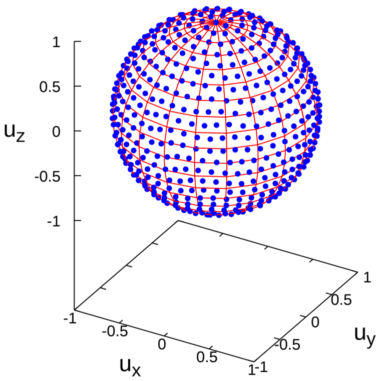


FIGURE 3.49: Representation of the $N_u = 801$ points over a surface of a unit sphere chosen for calculating the orientational averages.

The three-dimensional paracrystal is expressed by [43]

$$S_{i,\mathrm{PT}}(\mathbf{q}) \;=\; \prod_{\chi=a,b,c} \Re\left[\frac{1+F_{\chi,i}}{1-F_{\chi,i}} - \frac{2F_{\chi,i}(1-F_{\chi,i}^{N_{\chi,i}})}{N_{\chi,i}(1-F_{\chi,i})^2}\right] \tag{3.48}$$

where $F_{\chi,i} = e^{-\frac{1}{2}\mathbf{q}^{\mathrm{T}}\boldsymbol{\sigma}^2_{\chi,i}\mathbf{q}+i\mathbf{q}\cdot\boldsymbol{\chi}_i}$, $\boldsymbol{\sigma}^2_{\chi,i} = g^2_{\chi,i}(\boldsymbol{\chi}_i \otimes \boldsymbol{\chi}_i)$ is the variance matrix and $\mathbf{q}^{\mathrm{T}}$ the vector transpose of $\mathbf{q}$.

3.3.6 SAXS and WAXS results of microfibers in artificial sea water and fresh water

All SAXS and WAXS experimental data, measured at the SAXS beamline of the Elettra synchrotron over the samples of the 10 microfiber solutions, sampled in a regular grid of solar simulator radiation days, have been systematically and successfully analysed with the GENFIT software by using the models introduced in Section 3.3.5. Fitting

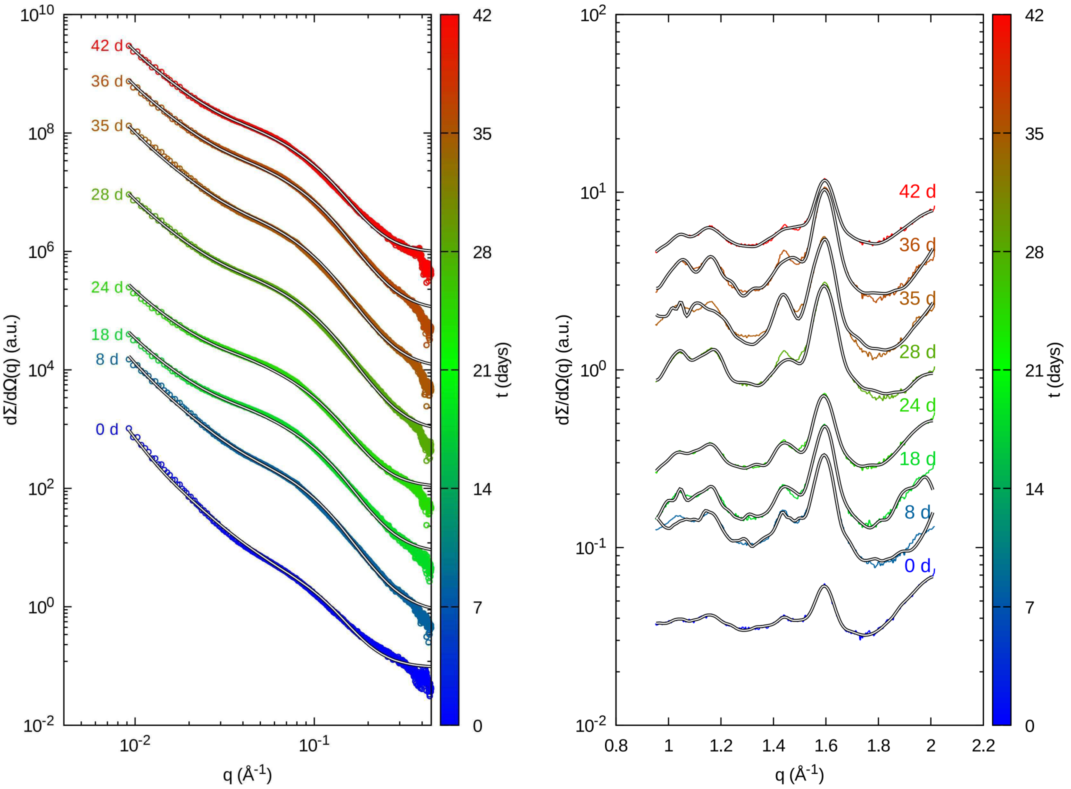


Figure 3.50: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **cotton** in **artificial sea water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

curves results are shown in Figs. 3.50, 3.53, 3.56, 3.58, 3.60, 3.63, 3.66, 3.68 and 3.70, 3.73. Fitting parameters of the SAXS curves are reported in Figs. 3.51, 3.54, 3.57, 3.59, 3.61, 3.64, 3.67, 3.69, 3.71 and 3.74. Fitting parameters of the WAXS curves that show diffraction peaks are reported in Figs 3.52, 3.55, 3.62, 3.65, 3.72 and 3.75.

Results are hereafter discussed, by separately considering each of the five fibers investigated and by comparing the results obtained in ASW and in fresh water.

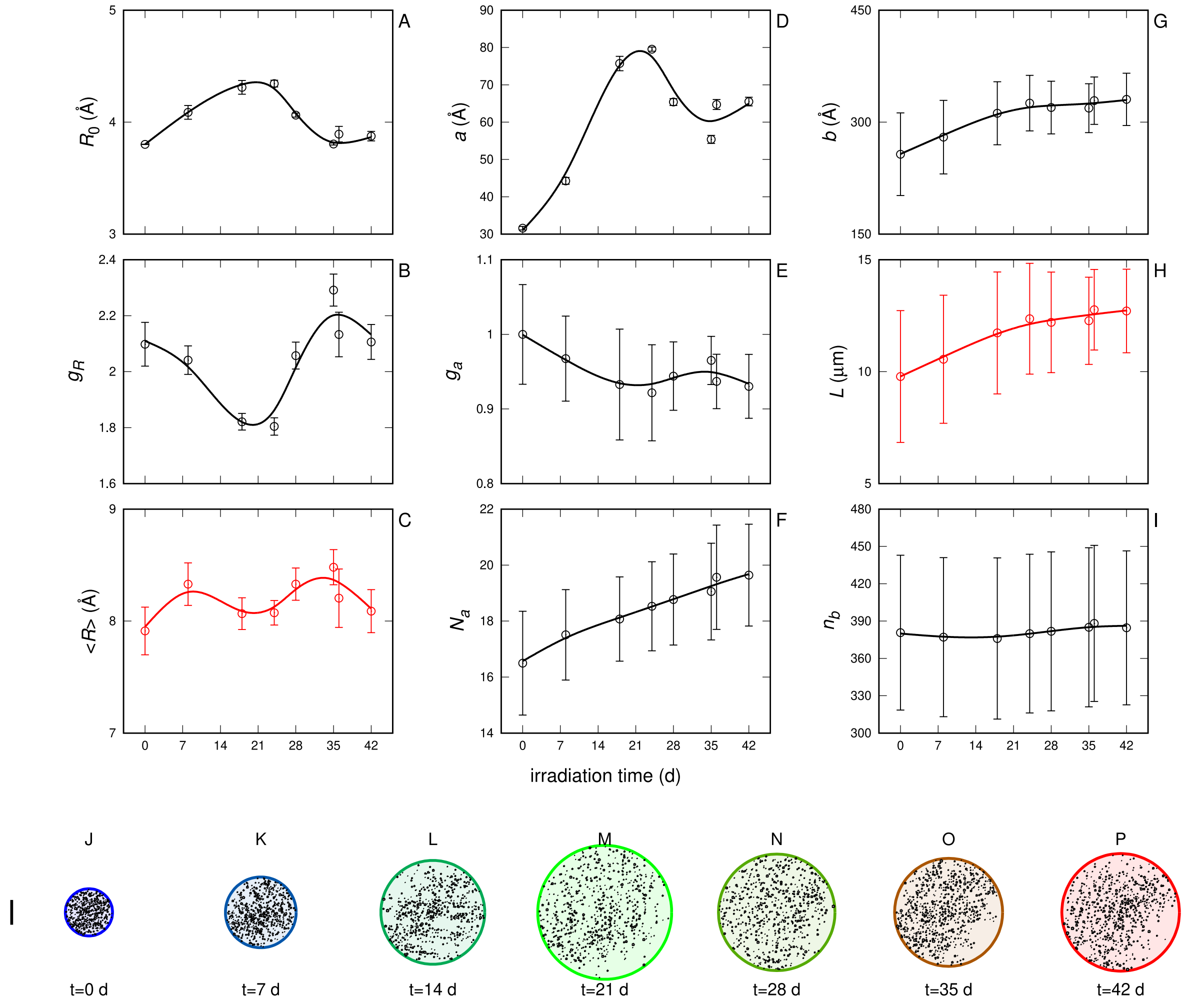


FIGURE 3.51: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **cotton** in **artificial sea water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

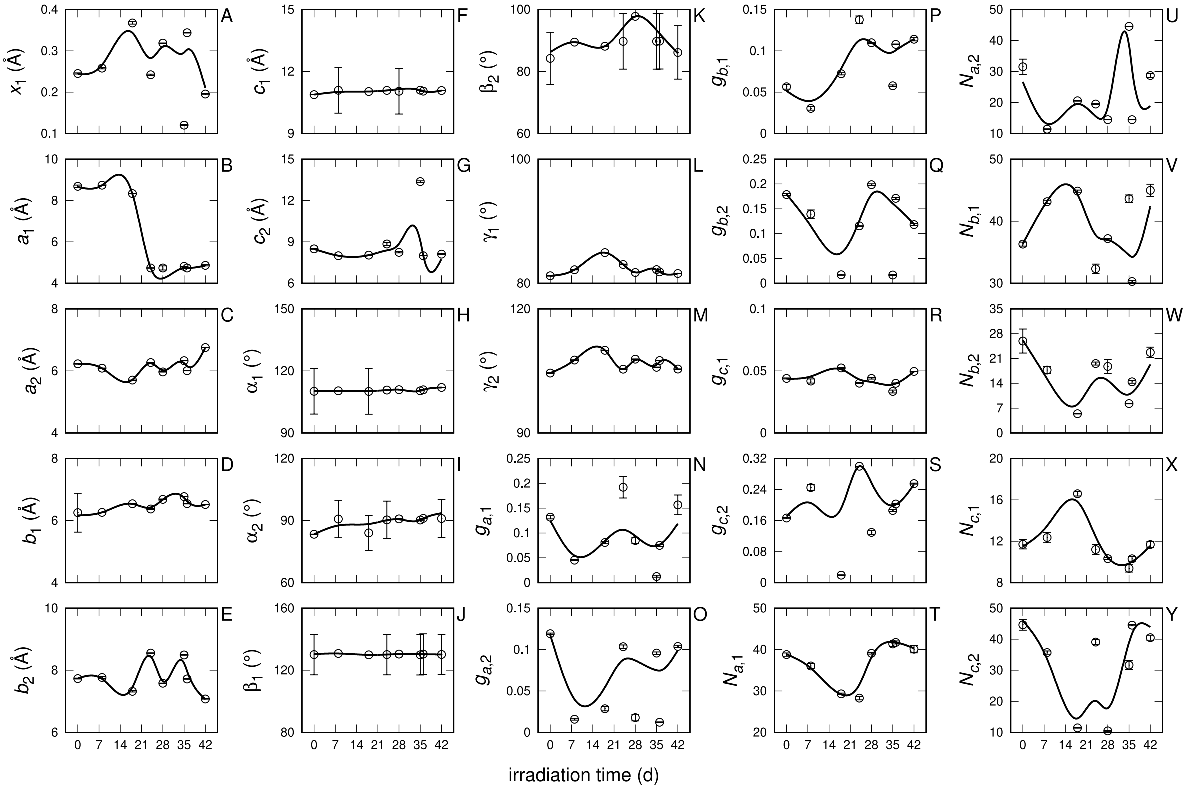


FIGURE 3.52: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **cotton** in **artificial sea water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

3.3.6.1 Cotton

SAXS results of irradiated cotton microfibers in artificial sea water (Figs. 3.50, left panel, and 3.51) show remarkable changes regarding the a parameter, which represents the average distance between microfibrils. a goes from 30 Å at the beginning of the irradiation time to a maximum of 80 Å around the 21st day, whereas N_a, whose square is the number of microfibrils embedded into the macrofibril, gradually increases underlying either the propensity of the biggest fibrils to fray or a possible incorporation of other macrofibrils together, which increases the number of microfibrils per macrofibril. On the contrary,

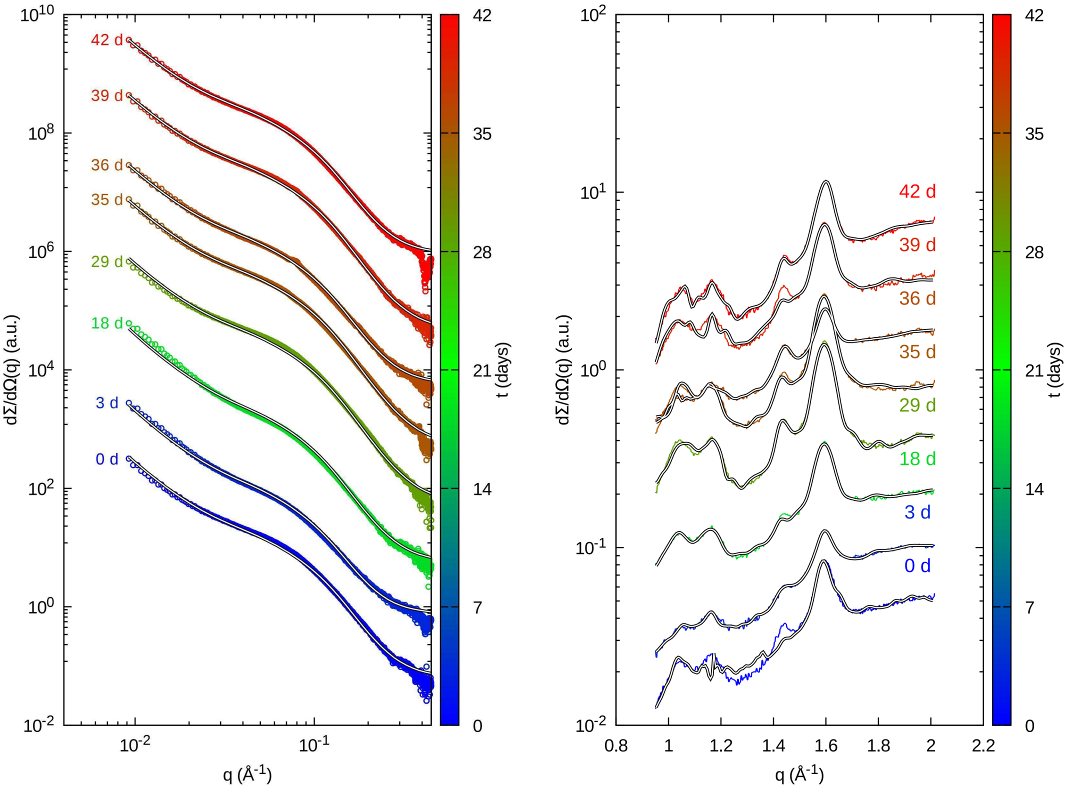


FIGURE 3.53: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **cotton** in **fresh water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

the microfibrils average radius R, as well as the length of the fully extended chain L, are kept constant. Basically, the irradiation of the solar simulator in artificial sea water brings the cotton microfibrils to get away from each other, and consequently it increases the macrofibril dimension that is a function of both a and N_a. This effect may also be provided by the permeation of water inside the macrofibril and the formation of repulsive forces that can widen the microfibrils distances.

In fresh water (Figs. 3.53, left panel, and 3.54) results are pretty different: the distance between microfibrils oscillates between 60 Å to 80 Å during the whole irradiation time,

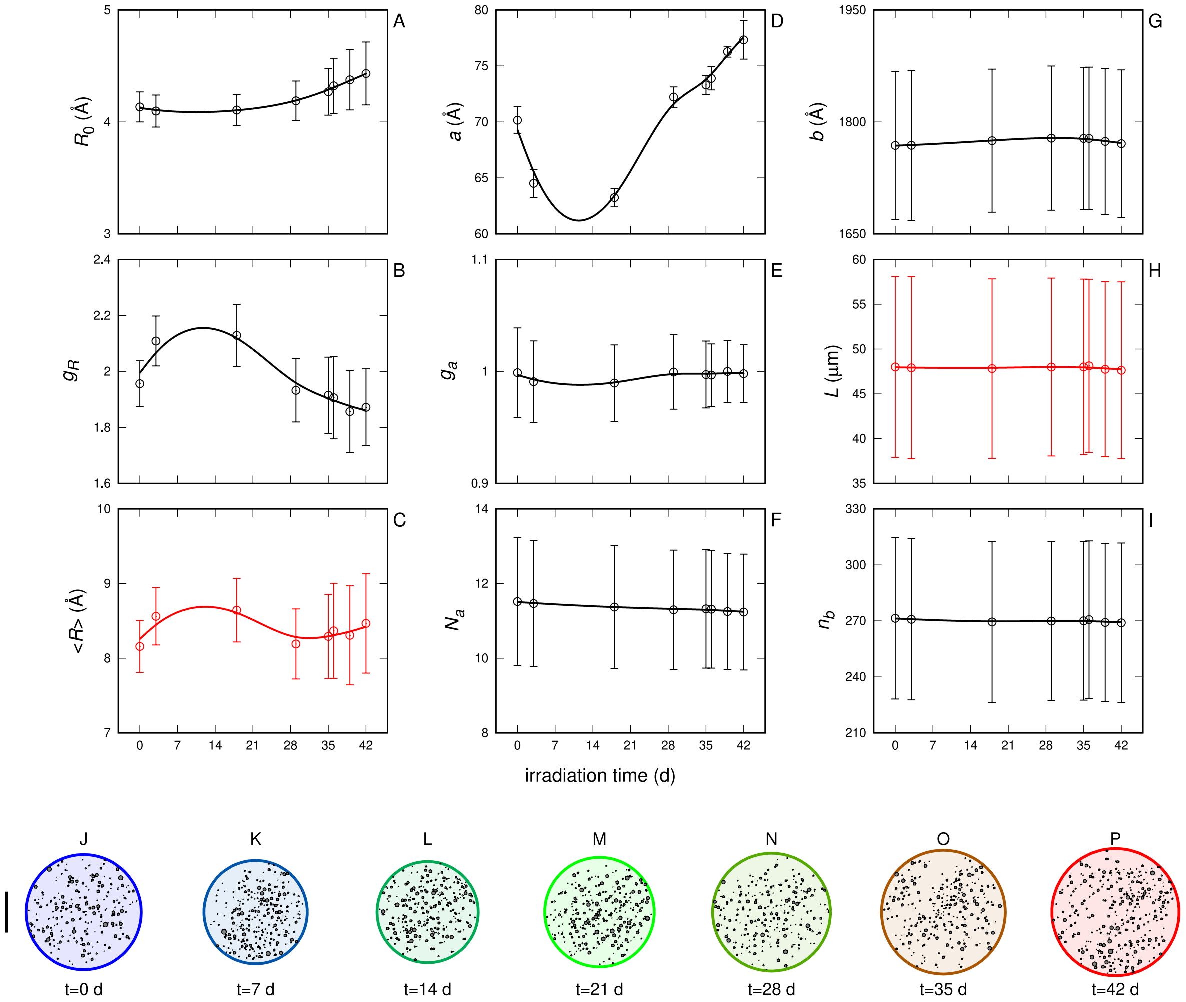


FIGURE 3.54: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **cotton** in **fresh water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

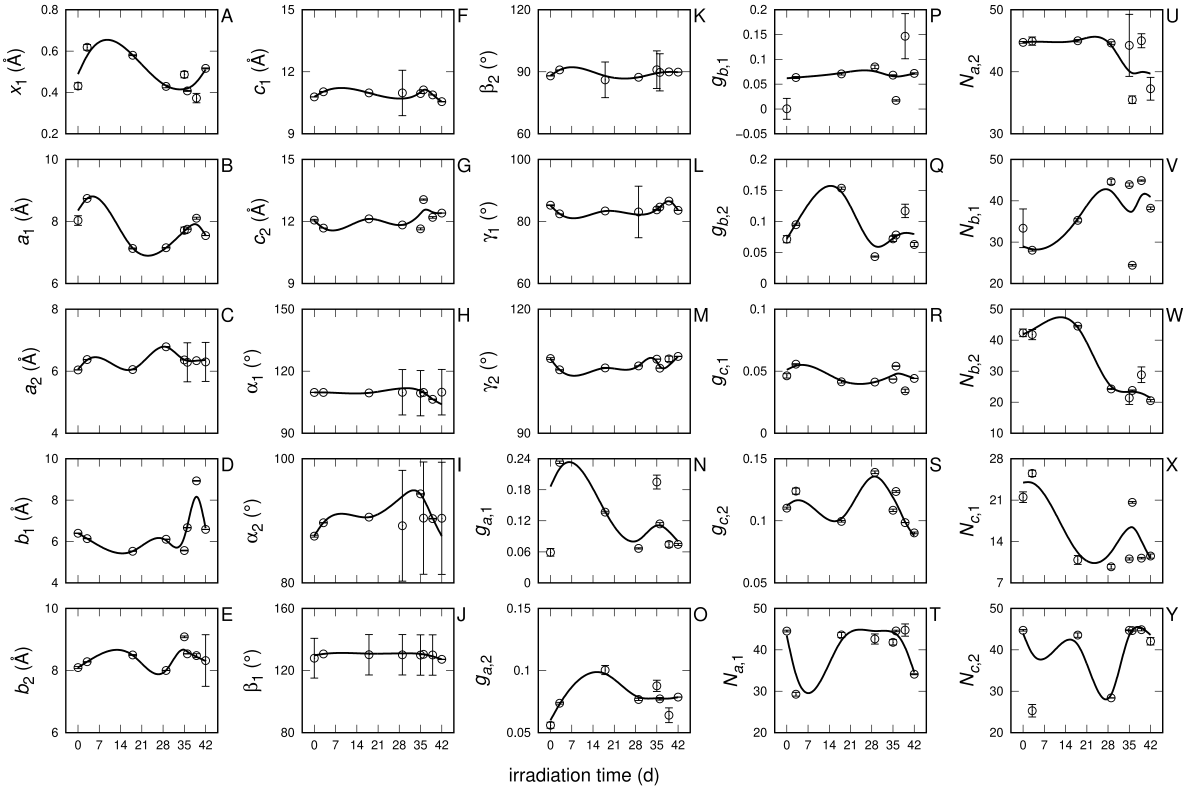


FIGURE 3.55: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **cotton** in **fresh water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

with no other remarkable changes in any other parameter, which means that fresh water does not have any particular effect on the degradation process of cotton microfibers.

WAXS results in artificial sea water are shown in Figs. 3.50, right panel, and fitting parameters are plotted as a function of the irradiation time in Fig. 3.52. WAXS results in fresh water are seen in Figs. 3.53, right panel, and parameters in Fig. 3.55. Fitting parameters confirm that cotton fiber are almost pure cellulose. In fact, according to literature [86, 50, 62], cellulose is formed by two distinct crystalline phases, one triclinic and the other monoclinic. Our fitting parameters show the presence of a phase 1 with

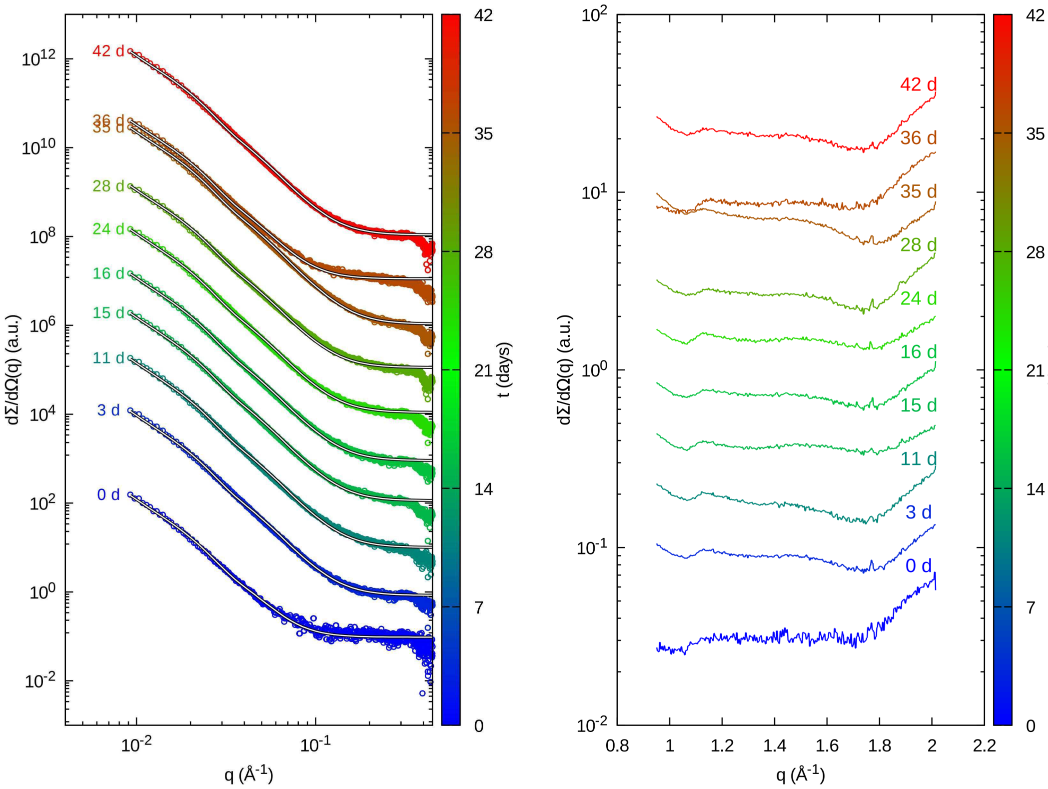


FIGURE 3.56: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **cellulose acetate** in **artificial sea water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

angles $\alpha_1 \approx 110°$, $\beta_1 \approx 130°$ and $\gamma_1 \approx 81°$ (panels H, J and L of both Figs. 3.52 and 3.55), very close to the one of the triclinic phase reported by [86] and a phase 2 with angles $\alpha_2 \approx 90°$, $\beta_2 \approx 90°$ and $\gamma_2 \approx 100°$ (panels I, K and M of both Figs. 3.52 and 3.55) that reflects a monoclinic phase. Interestingly, considering the parameter x_1, which gives the relative weight of crystalline phase 1 (panel A both Figs. 3.52 and 3.55) in respect to the two crystalline phases, in artificial see water the monoclinic phase is dominant, whereas in fresh water the two phases have a similar weight. Regarding the triclinic phase (phase 1) and considering the lattice lengths a_1, b_1 and c_1 (panels B, D and F of both Figs. 3.52 and 3.55), we note that, in ASW, a_1 changes from ≈ 8.5 Å to approx 4.5 Å

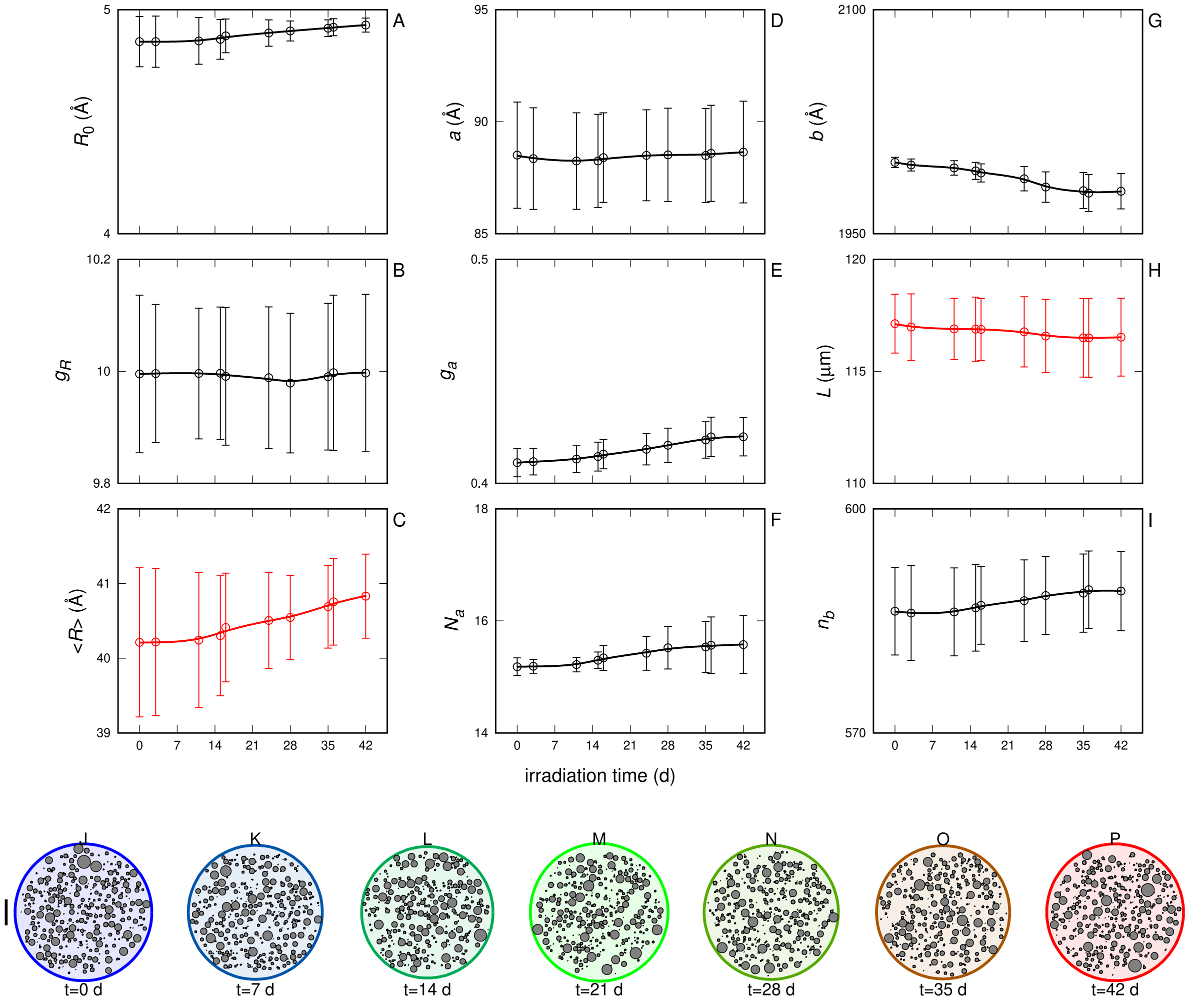


Figure 3.57: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **cellulose acetate** in **artificial sea water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

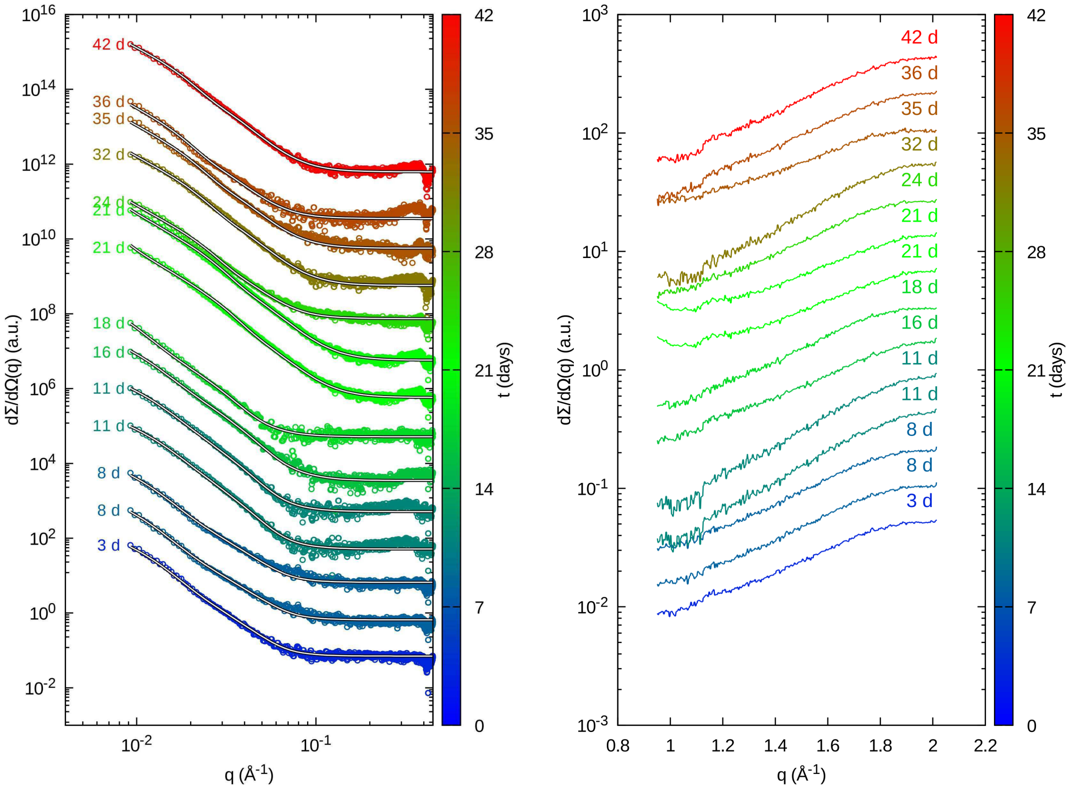


FIGURE 3.58: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **cellulose acetate** in **fresh water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

as a function of the irradiation time, whereas in fresh water it remains almost constant to 8.0 Å. On the contrary, the parameters b_1 and c_1, in both ASW and fresh water, are $\approx$ 6 Å and $\approx$ 10 Å, respectively, such us the values reported by Ref. [86] and are not greatly affected by the irradiation time. Regarding the lattice sides a_2, b_2 and c_2 of the phase 2 (monoclinic phase, panels C, E and G of both Figs. 3.55 and 3.55), we notice that for both kinds of water, their values are similar and quite constant with the irradiation time. These values are $a_2 \approx 6$ Å, $b_2 \approx 8$ Å and $c_2 \approx 12$ Å, very similar to the ones found by Ref. [86]. The other model parameters, in particular the unit cell sizes for both the triclinic and monoclinic phases, show variations with irradiation time that do not allow

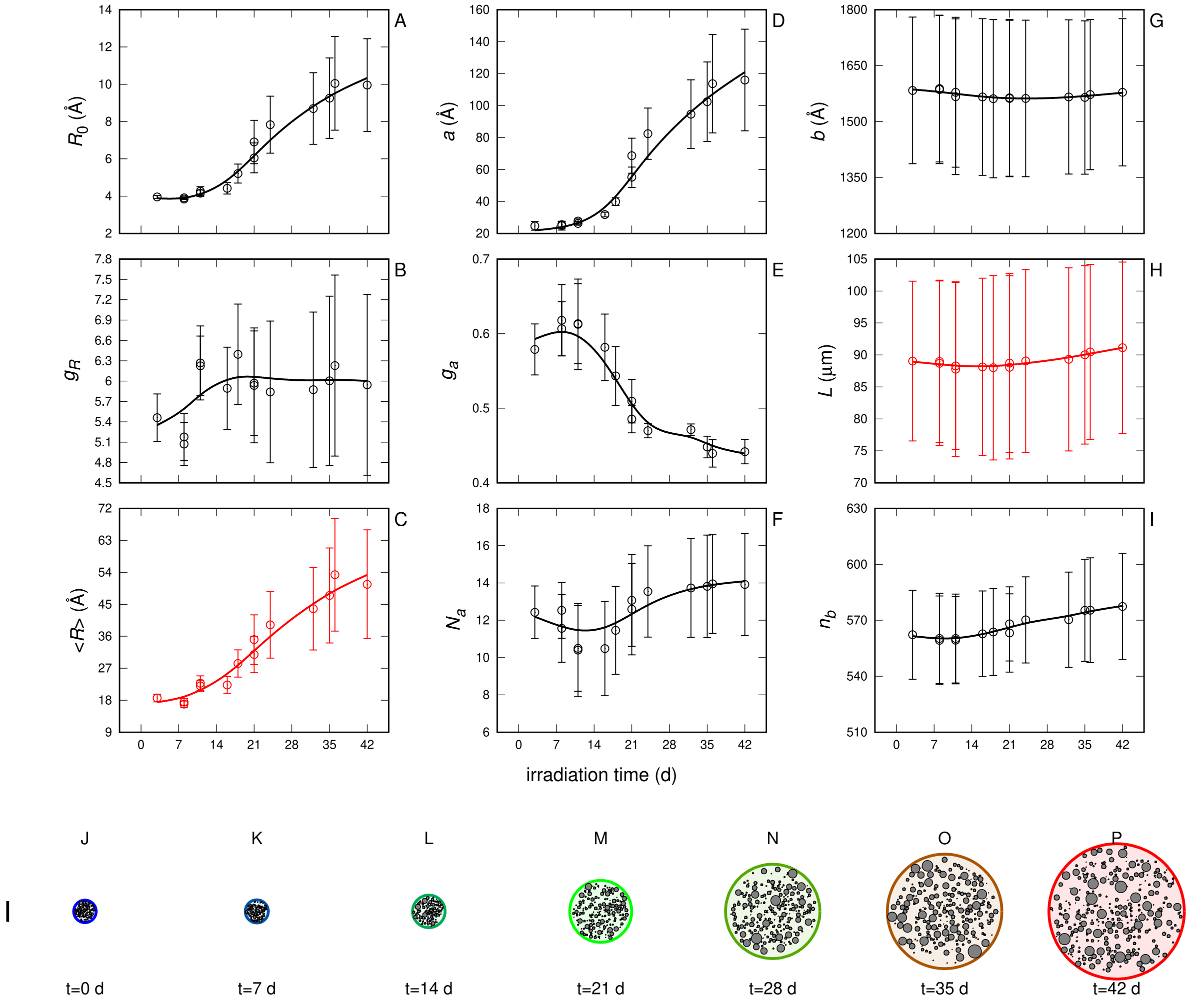


FIGURE 3.59: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **cellulose acetate** in **fresh water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

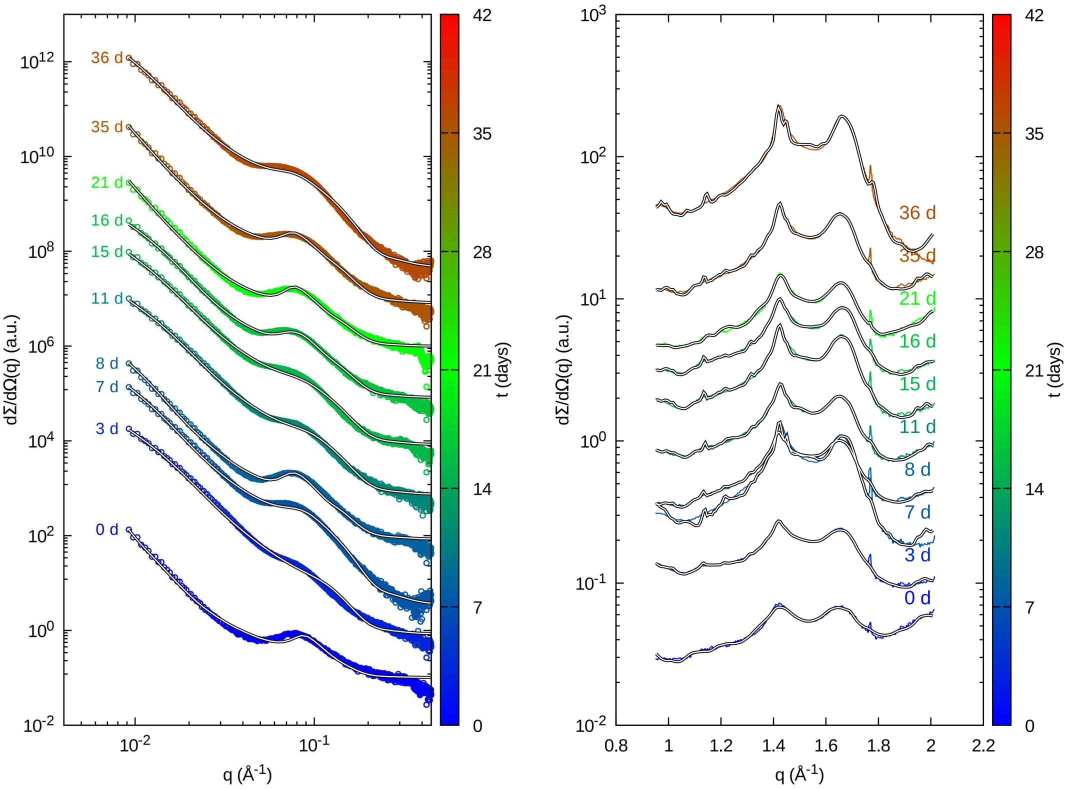


Figure 3.60: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **polyamide** in **artificial sea water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

a clear assessment of a UV effect, probably due to incorrect calibration of the WAXS data or heterogeneity of fibre sampling. We also note that the average crystallite size, which can be calculated from the product of cell size and cell number, is larger than the average microfibril radius determined from the SAXS curves. This result may indicate that the amorphous contribution, interpreted by the combination of a Gaussian and a power function (Eq. 3.46) and due to individual microfibrils seen by SAXS, is much larger than the crystalline contribution.

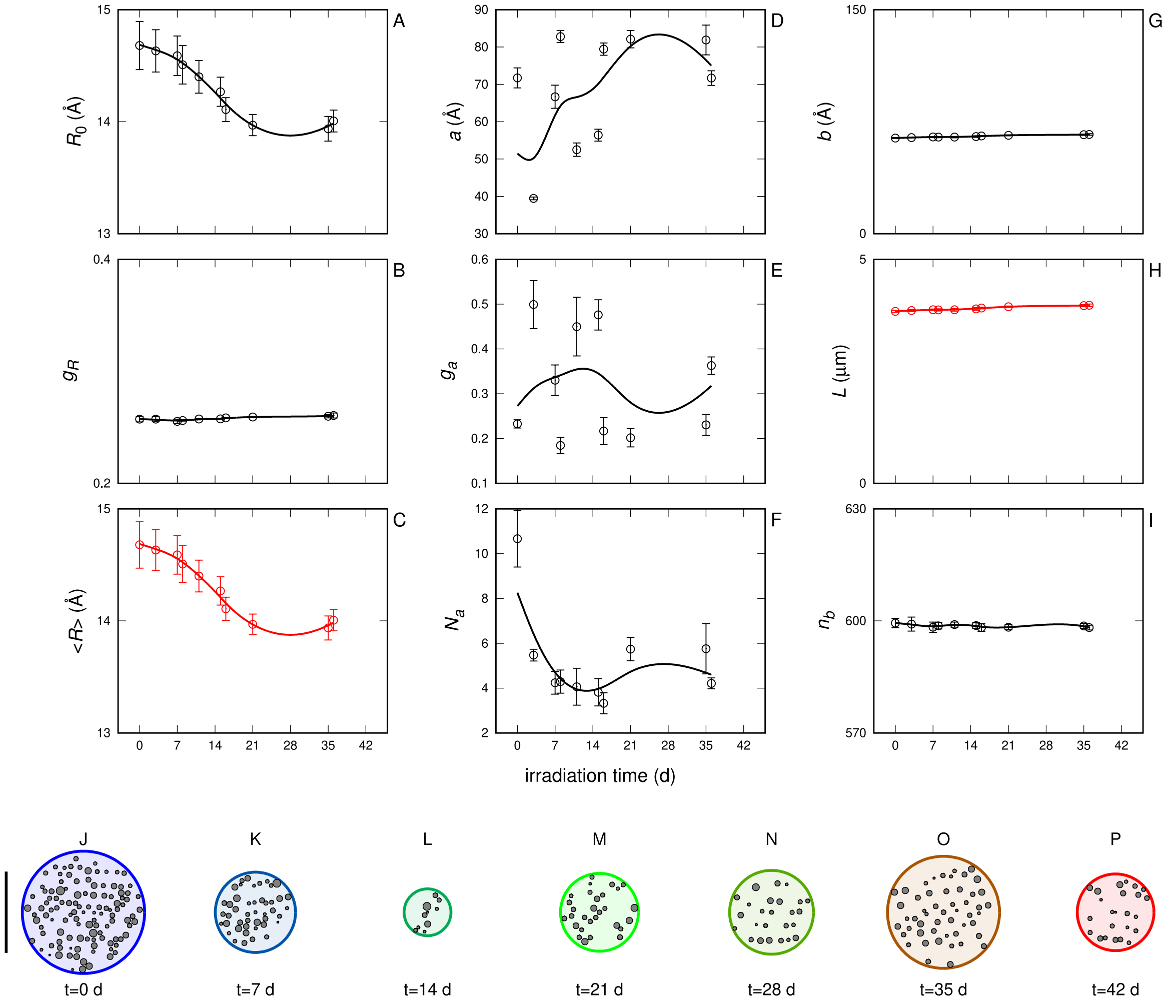


FIGURE 3.61: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **polyamide** in **artificial sea water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

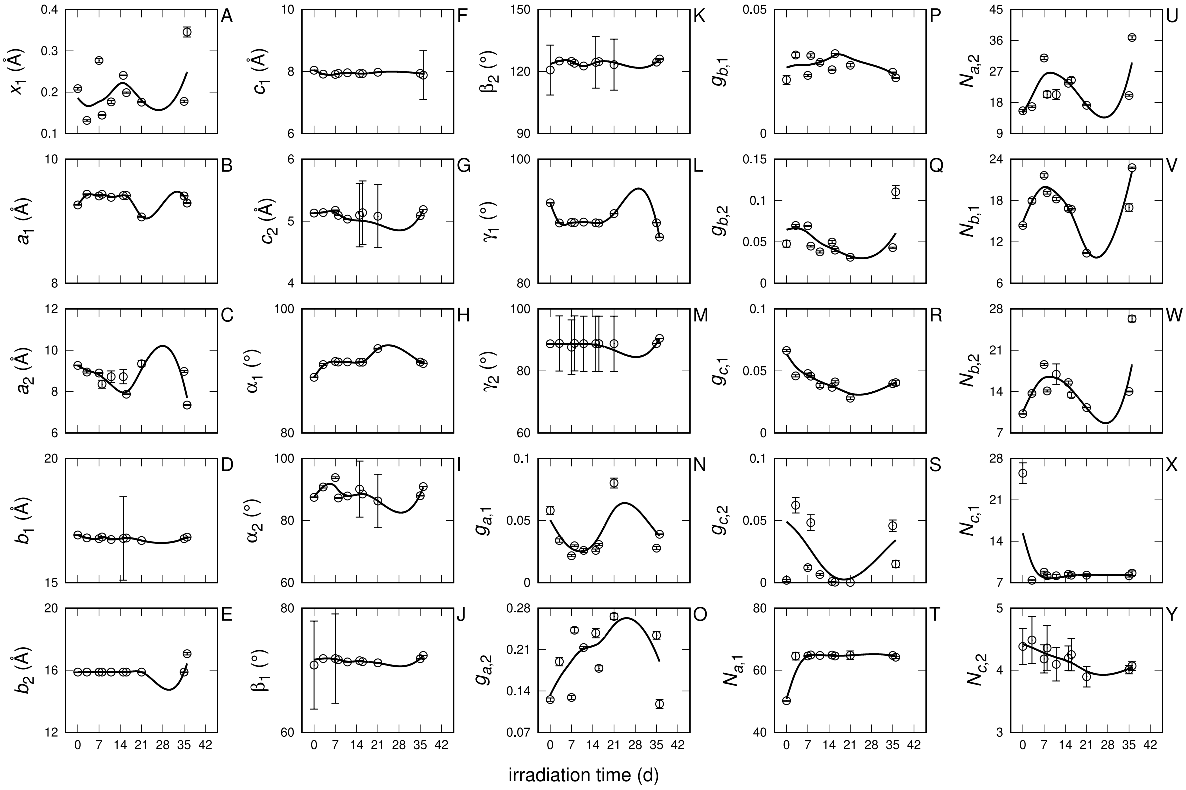


FIGURE 3.62: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **polyamide** in **artificial sea water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

3.3.6.2 Cellulose Acetate

SAXS results of microfibrils of cellulose acetate in artificial sea water (Figs. 3.56, left panel, and 3.59) have bigger initial diameters in respect to the other fibers analysed, but during the irradiation time they do not experience any changes in dimension or in length, and the same effect is seen for the whole macrofibril itself. Hence, the irradiation of cellulose acetate in artificial sea water seems to keep stable the microfiber structure, with no particular effects in their degradation process.

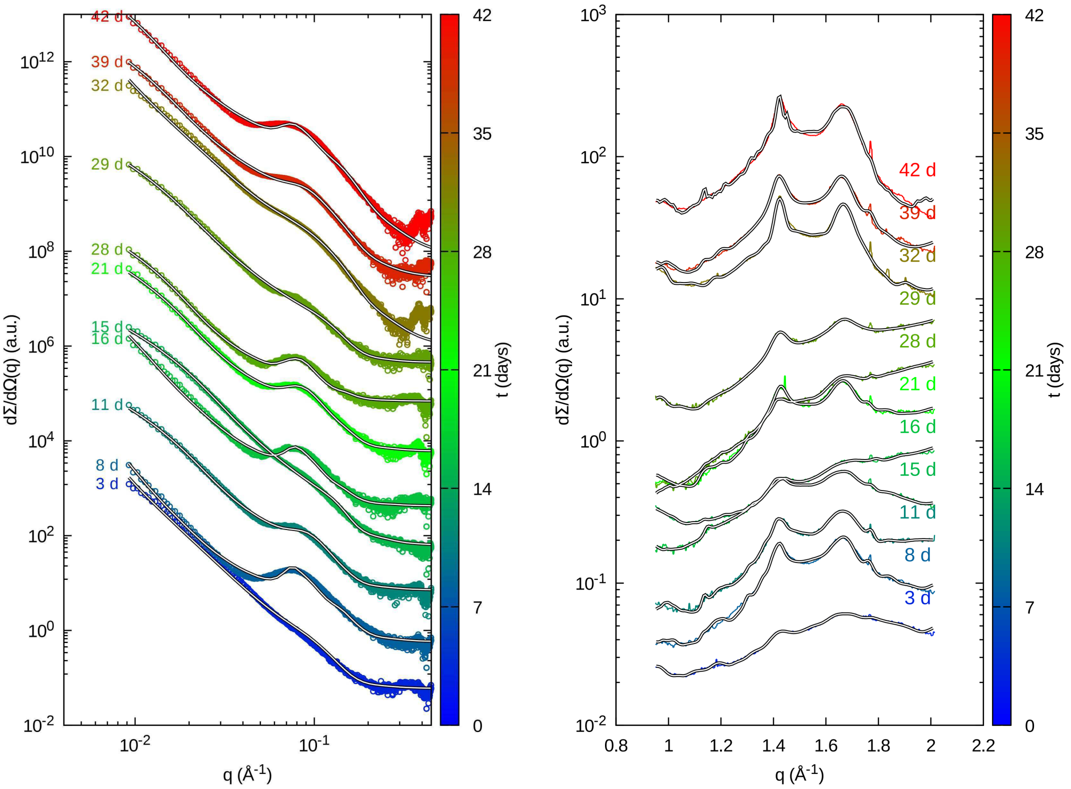


FIGURE 3.63: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **polyamide** in **fresh water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

An almost opposite phenomenon occurs when cellulose acetate is surrounded by fresh water (Figs. 3.58, left panel, and 3.59). The microfibrils average radius R varies from 18 Å to 50 Å as the irradiation time passes, and the distance between nanofibers a follows a similar trend, increasing the space between one fiber to another up to ≈ 100 Å. In this particular case, g_a is another important variable to take into account, because it is not as constant as it was in artificial sea water, but it decreases slowly when the irradiation time passes, showing a reduction in the distortion of the hexagonal lattice and therefore a more ordered fibers arrangement.

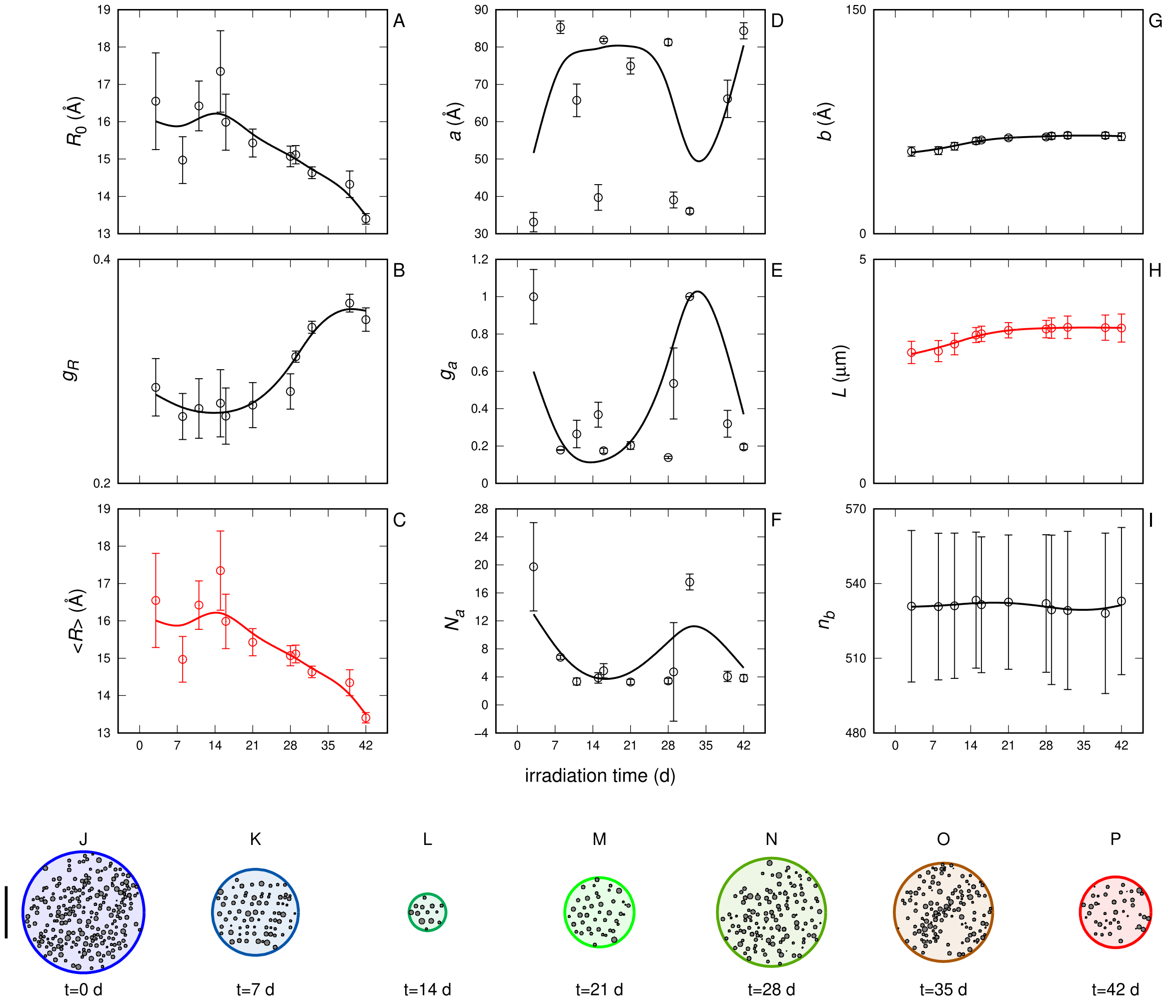


FIGURE 3.64: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **polyamide** in **fresh water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

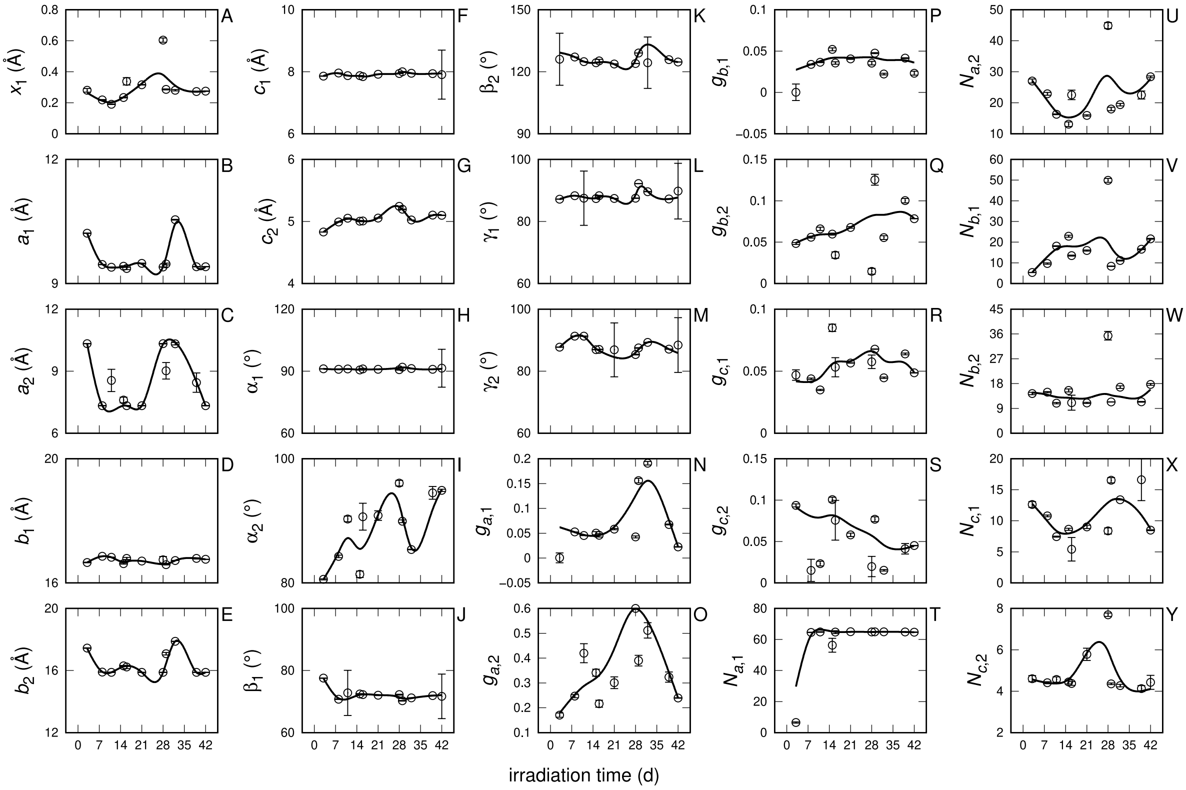


FIGURE 3.65: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **polyamide** in **fresh water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

3.3.6.3 Polyamide

Polyamide microfibers have been analysed with SAXS in artificial sea water (Figs. 3.60, left panel, and 3.61) and freshwater (Figs. 3.63, left panel, and 3.64), but in both cases data seem not to be such representative. This is probably due to the sampling activity that may have introduced too much variability in the analysis. The most reliable values are found in R, which undergoes a slight decrease when the irradiation time passes, and in L that shows the absence of fragmentation of polyamide during all the experiment, being the fiber length constant for all the process.

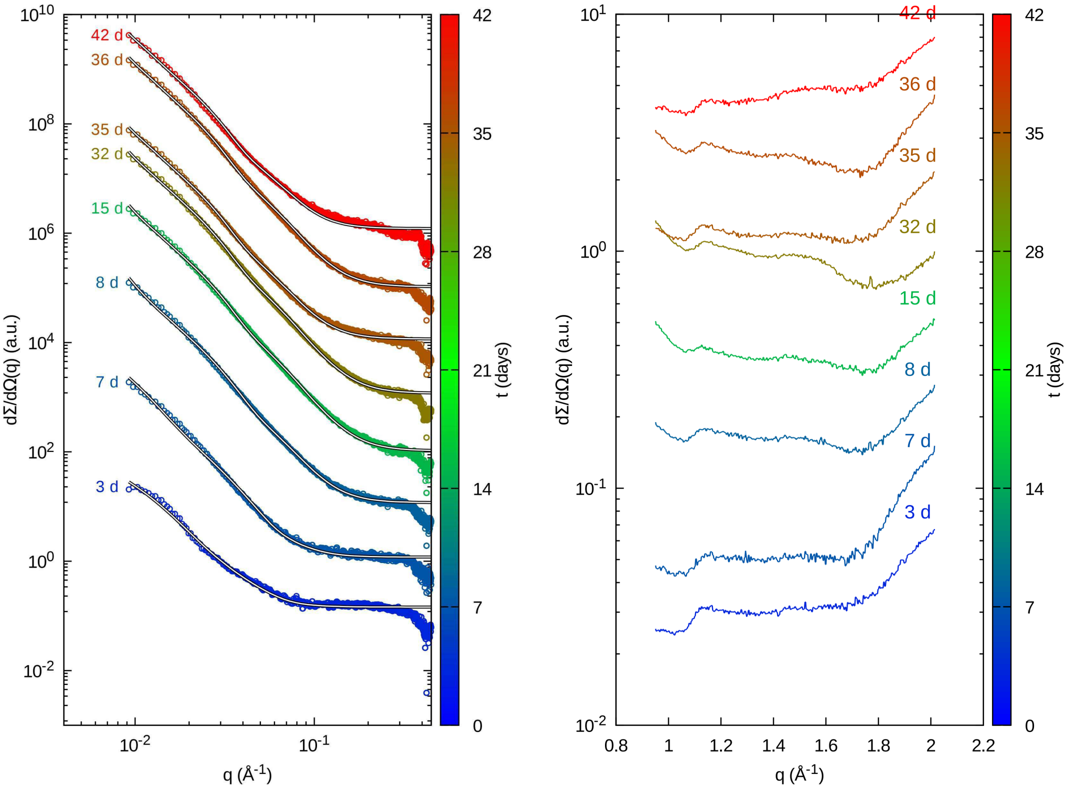


Figure 3.66: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **polyester** in **artificial sea water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

WAXS results in artificial sea water are shown in Figs. 3.60, right panel, and fitting parameters are plotted as a function of the irradiation time in Fig. 3.62. WAXS results in fresh water are seen in Figs. 3.63, right panel, and parameters in Fig. 3.65. The pattern of the WAXS curves, consisting of a broad central peak, attributed to the amorphous component, and two narrower peaks, representing two different crystalline orders, is quite similar to the results obtained from Ref. [18]. The two crystalline phases, both of monoclinic type, were named $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ (we use bold font in order to distinguish the phases' name by the angles). The excellent fit of the WAXS curves, obtained with Eq. 3.46, which exactly predicts the presence of an amorphous and two crystalline regions, allows

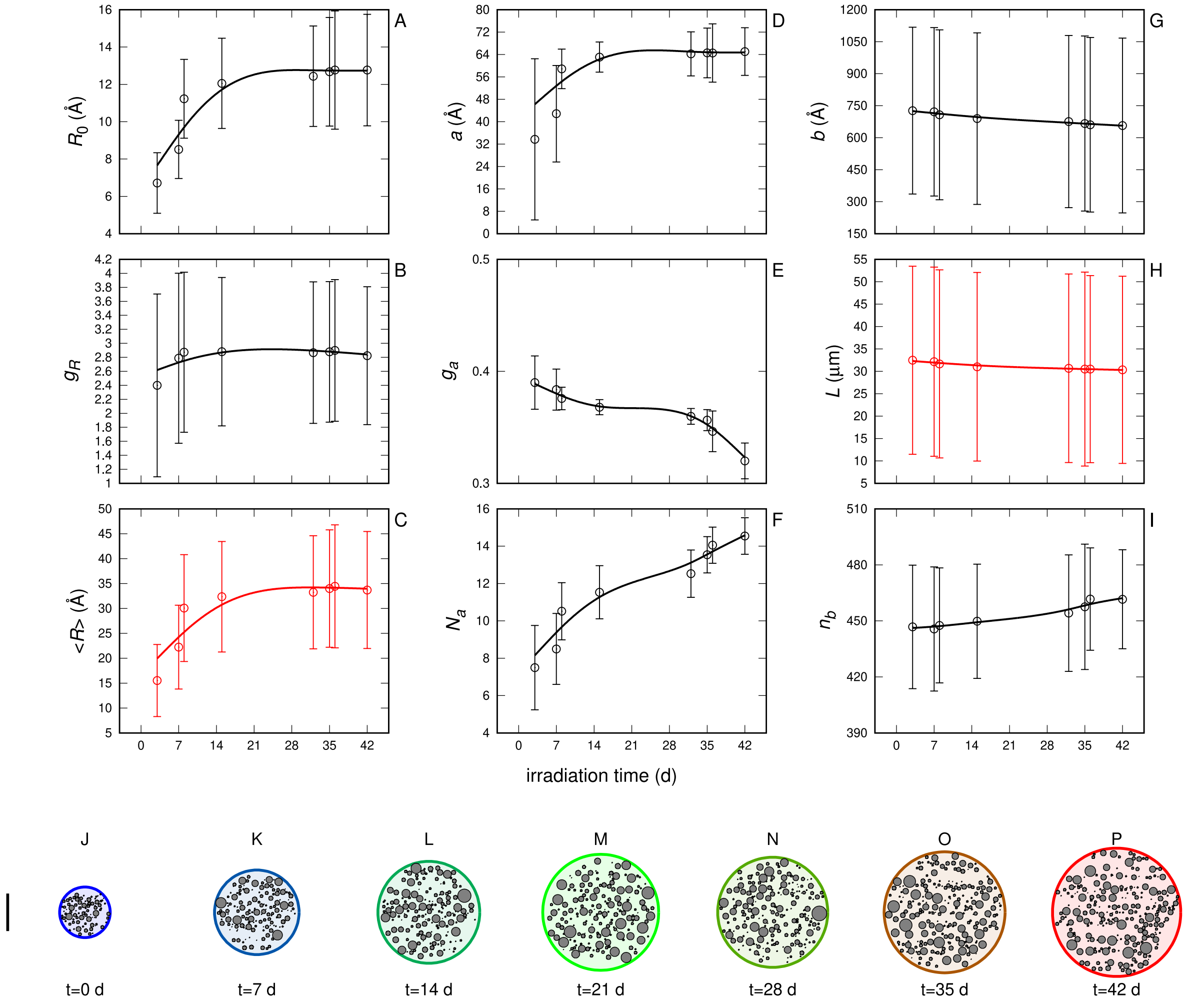


FIGURE 3.67: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **polyester** in **artificial sea water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

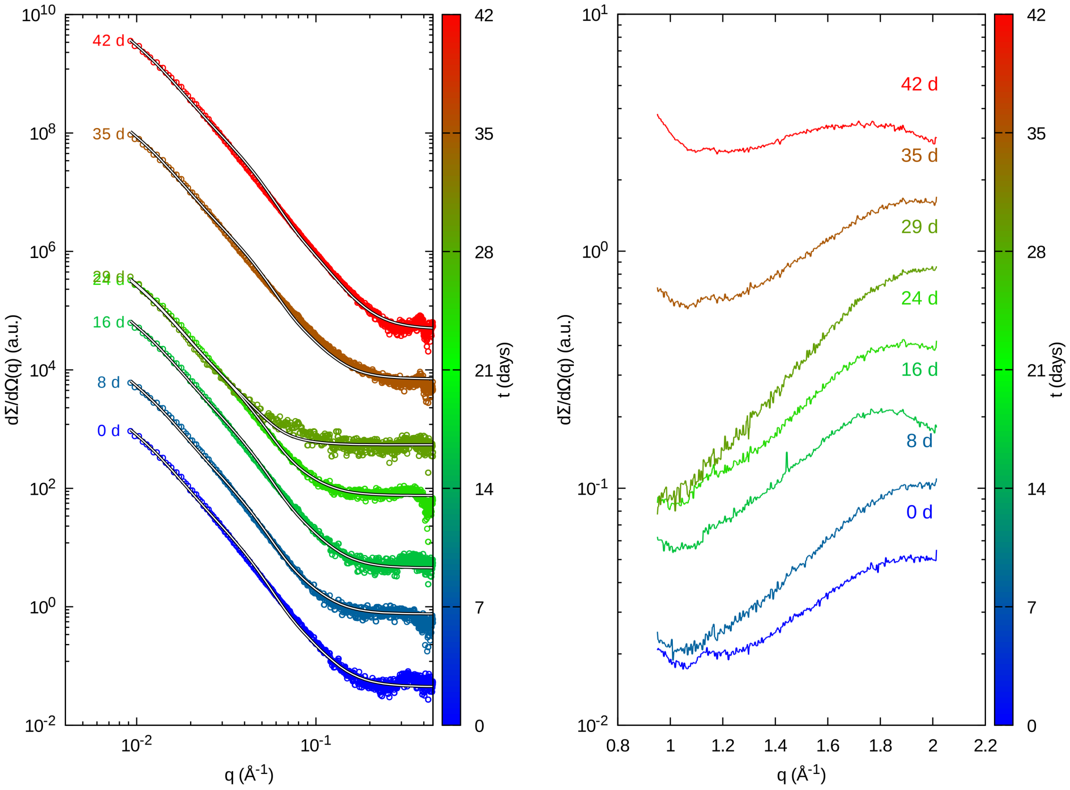


FIGURE 3.68: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **polyester** in **fresh water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

the crystallographic parameters of the two phases, shown in Figs. 3.62 and 3.65 as a function of the irradiation time, to be determined. It can first be seen that two of the angles of phase $\boldsymbol{\alpha}_1$ (α_1 and γ_1, panels H and L of Figs. 3.62 and 3.65, respectively), as well as two of the angles of phase $\boldsymbol{\alpha}_2$ (α_2 and γ_2, panels I and M of Figs. 3.62 and 3.65, respectively), are close to 90°, confirming that both phases are monoclinic. The third angle, for $\boldsymbol{\alpha}_1$ results $\beta_1 \approx 70°$ (panel J of Figs. 3.62 and 3.65), and for phase $\boldsymbol{\alpha}_2$ we have $\beta_2 \approx 125°$ (panel K of Figs. 3.62 and 3.65). It can be seen that neither of these two angles change as a function of irradiation time. Regarding the length of the cell parameters, for phase $\boldsymbol{\alpha}_1$, we have $a_1 \approx 9.2$ Å, $b_1 \approx 16.7$ Å and $c_1 \approx 7.9$ Å (panels B, D and F of

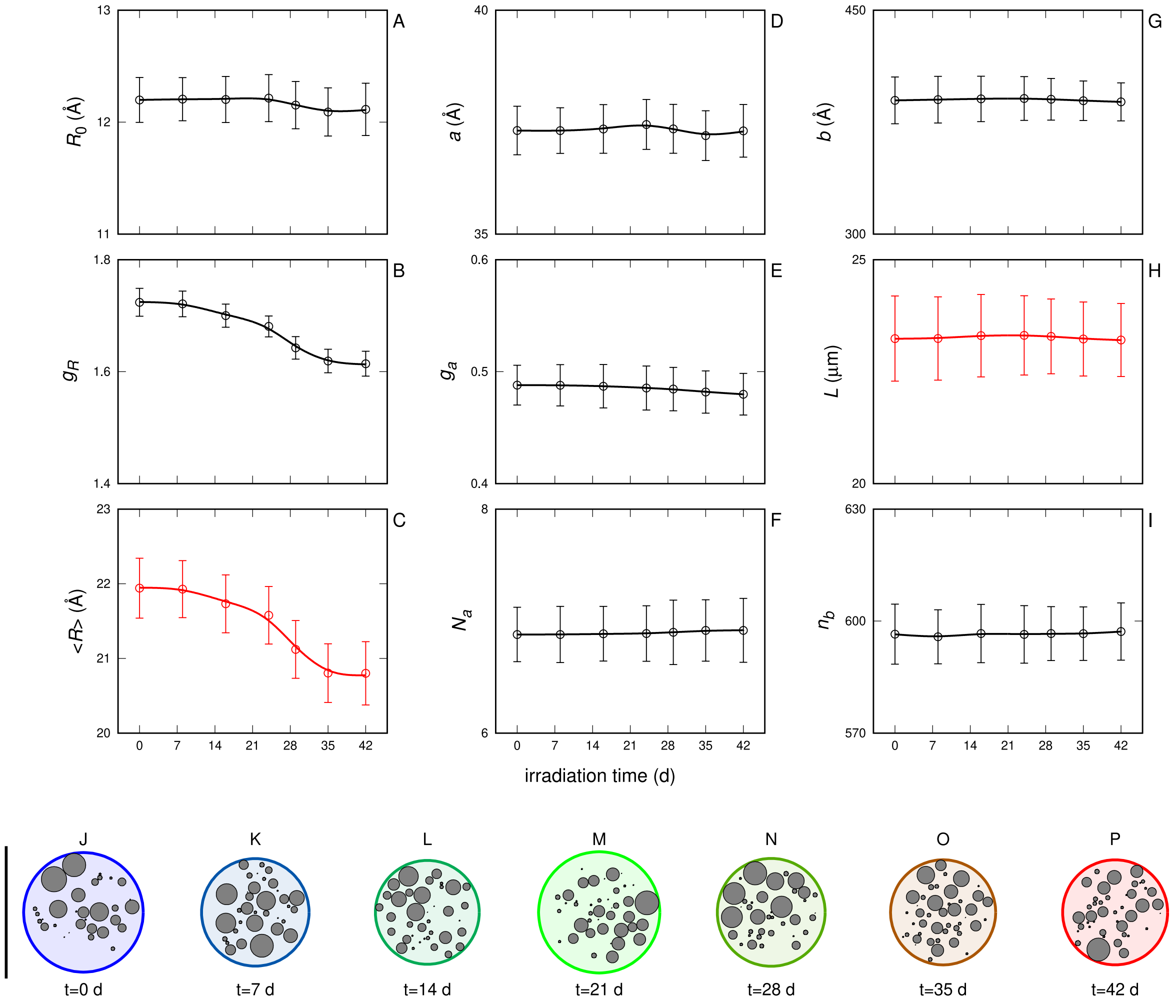


Figure 3.69: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **polyester** in **fresh water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a.

The vertical black line on the left of panel J represents the length of 500 Å.

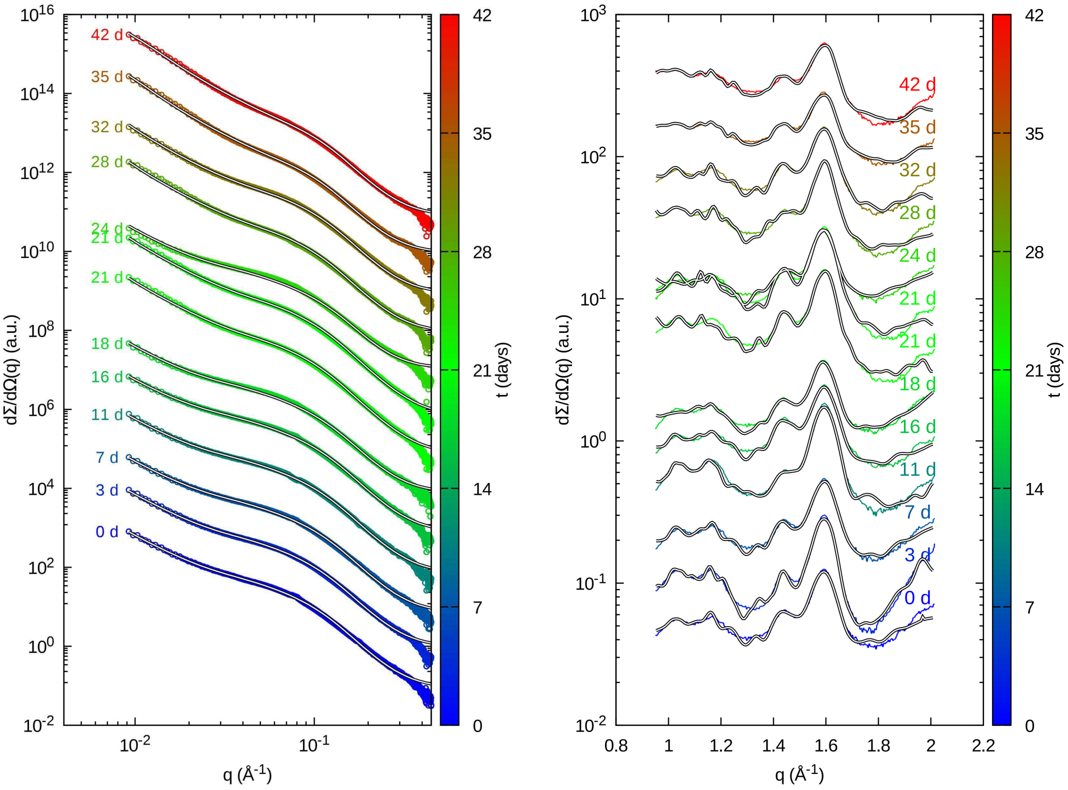


FIGURE 3.70: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **linen** in **artificial sea water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

Figs. 3.62 and 3.65, respectively), while for the $\boldsymbol{\alpha}_2$ phase we have $a_2 \approx 8$ Å, $b_2 \approx 15.9$ Å and $c_2 \approx 5.1$ Å (panels C, E and G of Figs. 3.62 and 3.65, respectively). None of these parameters show a significant change with irradiation time. No significant effect is also noticed when switching from artificial sea water to fresh water, as, on the other hand, can also be seen from a simple qualitative observation of the WAXS curves.

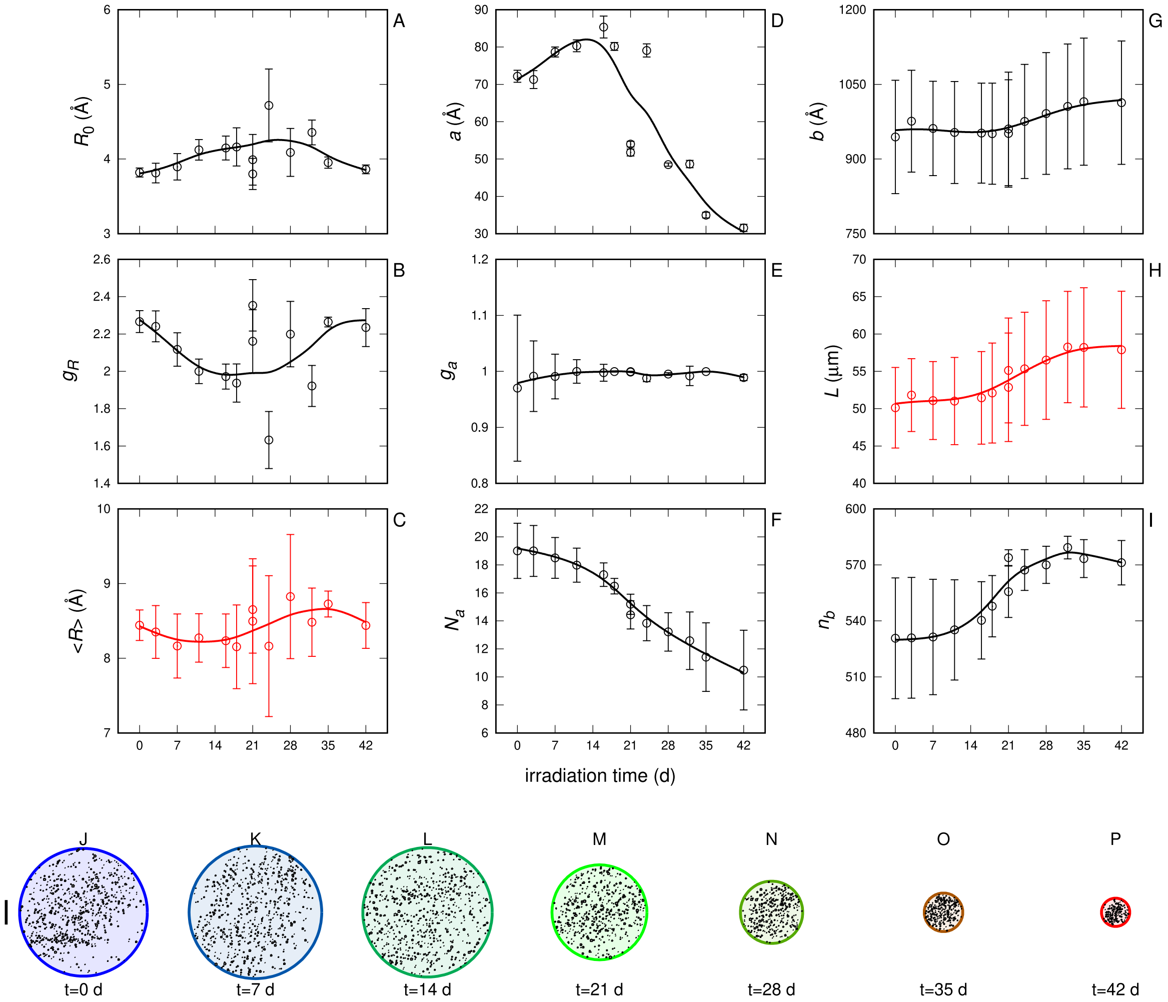


FIGURE 3.71: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **linen** in **artificial sea water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a. The vertical black line on the left of panel J represents the length of 500 Å.

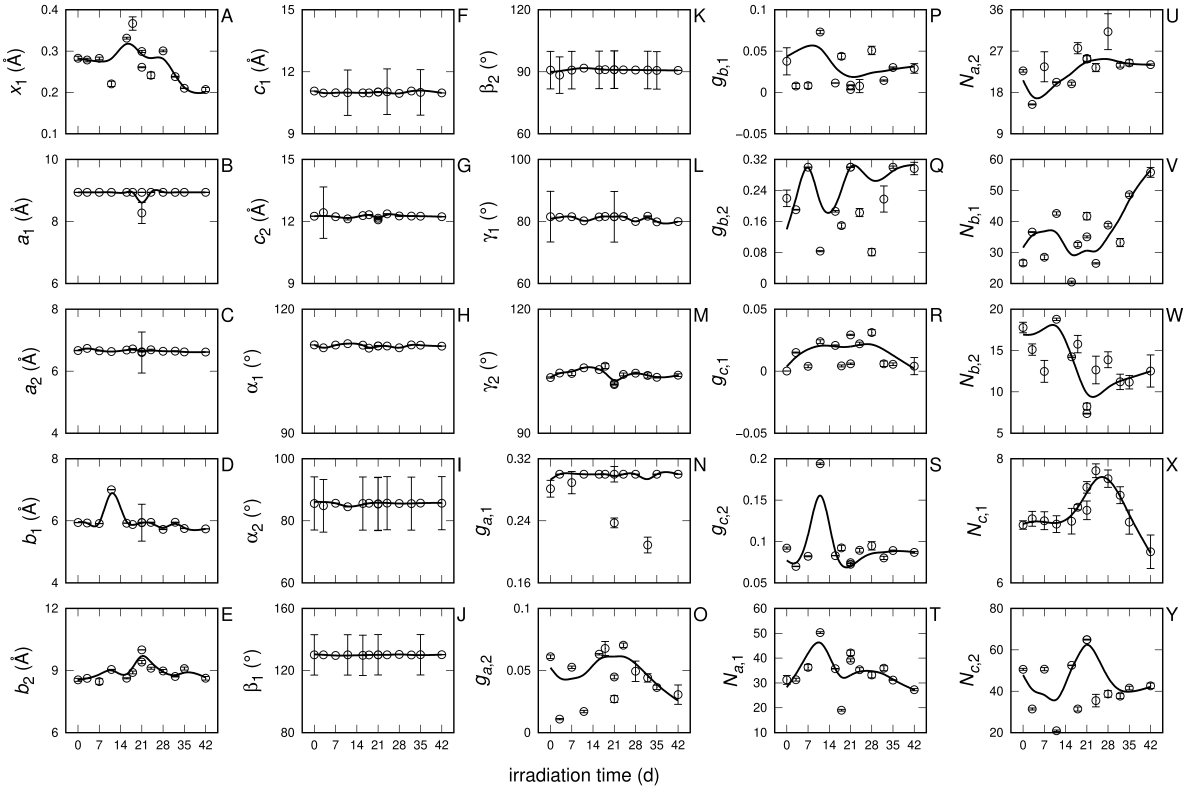


FIGURE 3.72: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **linen** in **artificial sea water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

3.3.6.4 Polyester

SAXS results of polyester in artificial sea water (Figs. 3.66, left panel, and 3.67) shows a behaviour similar to the one of cotton. Microfibrils radius starts to gradually increase until it reaches a plateau at around 20 Å from the third week of exposure. The parameter a shows a similar behaviour, meaning that the macrofibril diameter expands and also becomes richer in the number of microfibrils, as indicated from the N_a positive trend. The same observation is not seen for polyester in fresh water (Figs. 3.68, left panel, and 3.69), because in that case values seem to be kept stable during all the process, with no

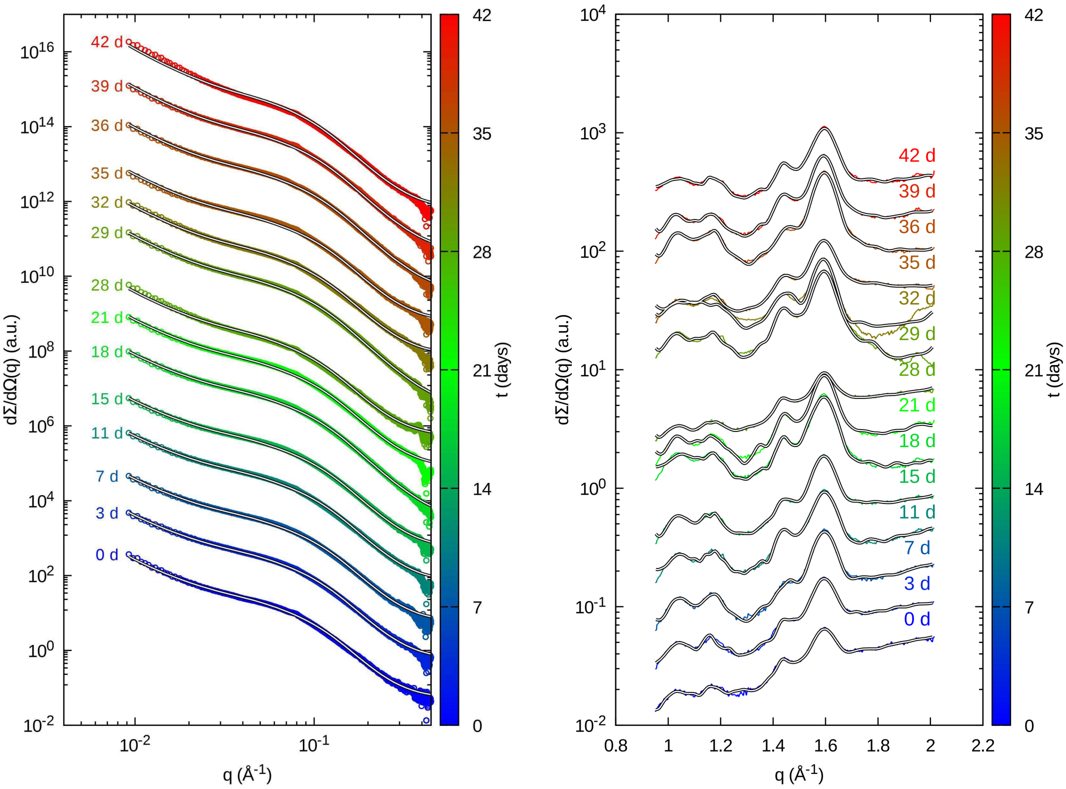


FIGURE 3.73: SAXS (left panel) and WAXS (right panel) curves for the microfibers of **linen** in **fresh water** recorded as a function of the irradiation time. The color of the points is selected according to vertical color box representing the the irradiation time. Solid black lines are the best fits obtained with GENFIT.

particular effect reported as the time passes, a typical signal of the absence of a effective degradation process.

3.3.6.5 Linen

Linen in artificial sea water, investigated with SAXS (Figs. 3.70, left panel, and 3.71), shows a different behaviour in respect to all the other fibers. In fact this is the only material that shows a decrease in size as the irradiation time passes. R is kept constant during all the irradiation process, and a decreases together with N_a, meaning that the

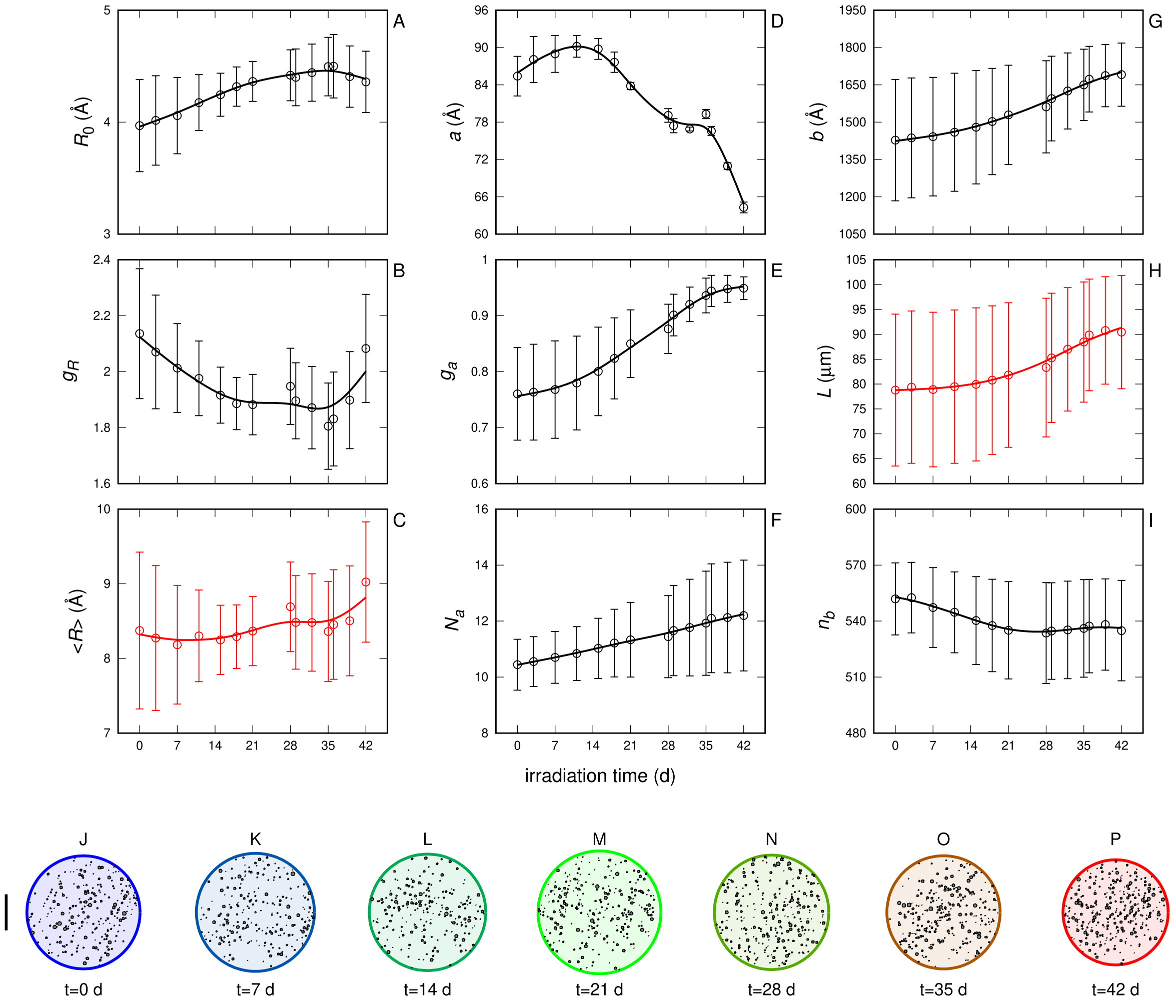


FIGURE 3.74: Panels A-I: fitting parameters (black points) and derived parameters (red points) of the SAXS curves for the microfibers of **linen** in **fresh water** reported as a function of the irradiation time. Smooth black or red curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters. Panels J-P: representation of the average microfiber cross section according to the fitting parameters. The small gray circles, representing the nanofibers, have been drawn with a radius sampled by the fitted Gaussian poly-dispersed distribution and in a position sampled according to the bidimensional para-crystal fitting parameters a, g_a and N_a.

The vertical black line on the left of panel J represents the length of 500 Å.

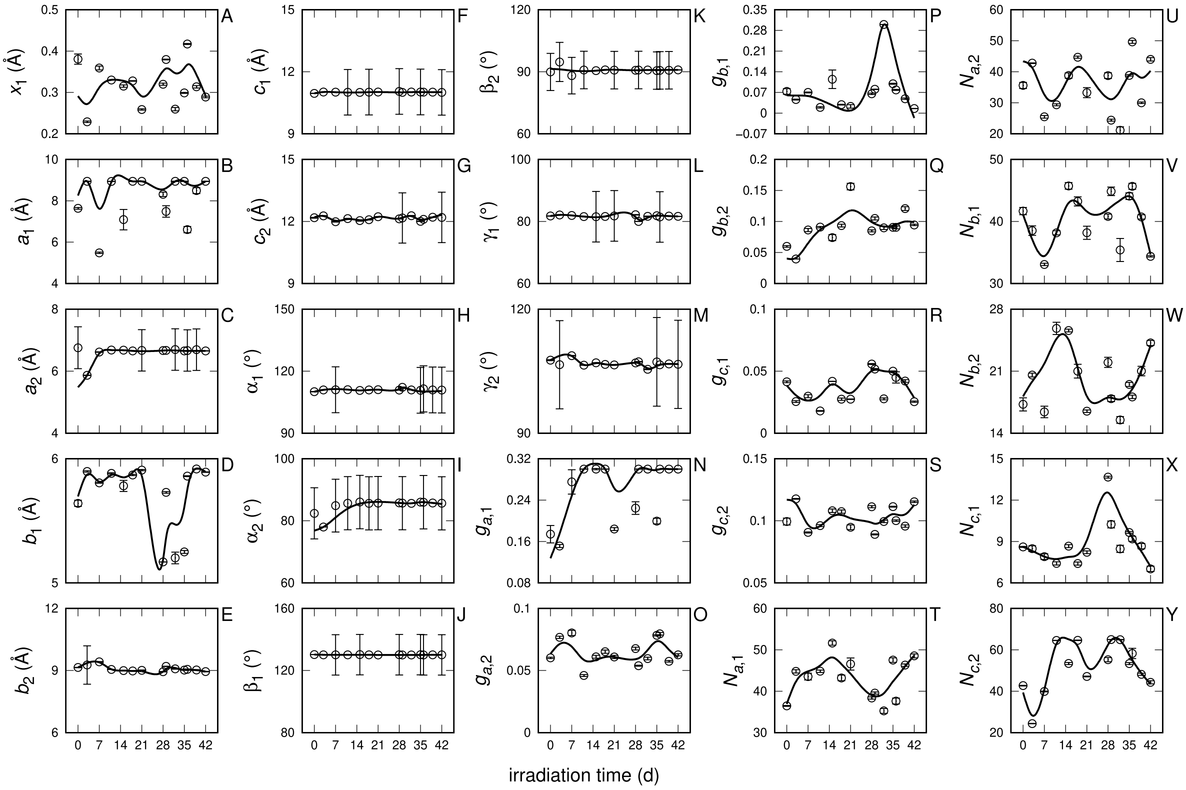


FIGURE 3.75: Panels A-Y: fitting parameters of the WAXS curves for the microfibers of **linen** in **fresh water** reported as a function of the irradiation time. Smooth black curves among the points have been obtained with cubic splines weighted with uncertainties of the parameters.

macrofibril might separate into smaller macrofibrils composed by a less amount of microfibrils, suggesting a shrinking of the macrofibril. In freshwater (Figs. 3.73, left panel, and 3.74) some features above mentioned show the same trend. In particular this behaviour concerns the parameters L, $R$and a, which slightly decreases even if they will not become as low as the linen fiber in artificial sea water. What is completely different in that case is the paramter N_a, which suggests that the number of microfibrils in respect to each lattice hexagonal direction increases with more days of irradiation.

WAXS results in artificial sea water are shown in Figs. 3.70, right panel, and fitting parameters are plotted as a function of the irradiation time in Fig. 3.72. WAXS results in fresh water are seen in Figs. 3.73, right panel, and parameters in Fig. 3.75. It is known that linen consists of a combination of amorphous cellulose and crystalline cellulose [39]. Crystalline cellulose, in turn, is present in two phases, called $I\alpha$ and $I\beta$, which differ in the patterns of hydrogen bonds [31]. From a qualitative point of view, the experimental WAXS curves are completely similar to those shown in Fig. 3 of Ref. [39], which concerns a study of ancient Egyptian linen. The interpretation model that we have adopted for the WAXS measurements is therefore also very well suited to describing the WAXS data of linen, collected as a function of the irradiation time. We can, in fact, evaluate the quality of the best fit curves shown in Figs 3.70 and 3.73. The fit parameters, shown in Figs. 3.72 and 3.75, highlight the existence of a triclinic phase, known in the literature as the $I\alpha$ phase, and a monoclinic phase ($I\beta$). In fact, the angles of the unit cell, for the phase $I\alpha$, turned out to be $\alpha_1 \approx 111°$, $\beta_1 \approx 130°$ and $\gamma_1 \approx 81°$ (panels H, J and L of both Figs. 3.72 and 3.75), in practice equal to the one obtained from the WAXS analysis of cotton, as well as similar to the one reported in Ref. [31]. For the $I\beta$ phase, we obtained the following angles: $\alpha_2 \approx 90°$, $\beta_2 \approx 90°$ and $\gamma_2 \approx 103°$ (panels I, K and M of both Figs. 3.72 and 3.75) also similar to those of the cotton and in agreement with literature. Concerning the the cell parameters' lengths, we got, for the $I\alpha$ phase, $a_1 \approx 8.9$ Å, $b_1 \approx 5.9$ Å and $c_1 \approx 11$ Å (panels B, D and F of Figs. 3.72 and 3.75, respectively) and, for the $I\beta$ phase, $a_2 \approx 6.6$ Å, $b_2 \approx 8.5$ Å and $c_2 \approx 12.2$ Å (panels C, E and G of Figs. 3.72 and 3.75, respectively). No significant differences of these parameters are found on going from artificial sea water to fresh water. In both cases, the relative weight of the $I\alpha$ phase, measured by the parameter x_1 (panel A of Figs. 3.72 and 3.75), is between 20 and 30%. However, in artificial sea water, despite large oscillations on x_1 are seen, it seems that the $I\alpha$ phase content decreases with increasing UV exposure.

3.3.7 Textile microfibers discussion and conclusions

Textile micro and nanofibers are emerging pollutants of great concern nowadays that need to be studied not only in respect to the toxicological effect that they might have on the organisms that absorb them, but also considering the mechanical and physical processes that lead to their generation. The simulation of the natural conditions present in the environment allowed us to undergo our samples to the equivalent of one year of sun exposure and the advanced SAXS analysis gave us information regarding the internal structure of the microfibers themselves. In some cases, the behaviour of the fibers seem to resemble each other, which is the case of cotton and polyester in artificial sea water, where the increasing of the irradiation time has brought to an enlargement in the macrofibril diameter. However, in other cases what happens is completely the opposite. For example, in the case of linen, the macrofibril dimension instead of increasing starts to shrinken. The changes in the structure might be ascribed to a possible radicalic reaction due to UV exposure. UV can form free radicals in the double bonds of polymers and then in the presence of the molecular oxygen dissolved in water they can trigger a series of processes that lead to further degradation. Besides, the presence of salt in water might also bring to differences in the effects of free radicals in the polymeric chain. Salts, in fact, are rich in electrons that can have a different interaction with the radicals, which can explain why there are differences between fibers in fresh and sea water. This work was our first overview on the mechanisms of degradation of textile microfibers in marine and fresh water: additional studies will be needed in order to better appreciate the contribution of each variable to the process and to have a clear insight in the degradation mechanism that takes place.

Chapter 4

General conclusions

In this PhD thesis, we have confirmed that the SAXS technique can be successfully applied to different research fields, starting from Biology and ending with Environmental Sciences. The versatility of SAXS is enhanced by the use of advanced thermodynamic and/or structural models. In fact the fitting of our data with suitable and tailored model had allowed us to obtain several kinds of information that contributes to generate a detailed insight in the most disparate fields. SAXS can be carried out alone or in combination with other techniques such as SEC, as I performed in my first studied system, which allowed overcoming the common problem of protein aggregation and obtaining information regarding the monomeric state of the translation elongation factor. In this case, in absence of the protein crystallized structure, a simple probability calculation was done to hypothesise which organisms owned the structure of the translation elongation factor more similar to the protein derived from *S. solfataricus*. The same logic was used also for the SAXS investigation of the complex between the protein and the enzyme aDHS, but in this case the aim of our analysis was understanding the stoichiometry of the complex and comparing the result with the recently crystallized complex structure. In this PhD thesis it was also shown how the combination of structural and thermodynamic characteristics of proteins allowed to get information about the stabilizing effects of cosolvents in solution. In particular, we described not only the preferential state of the protein in solution, but also the favoured water solvation properties on the protein surface as a function of the protein concentration, the temperature, the modified sugars concentration and also the proteins and sugars type. Then, last but not least, the SAXS analysis was also applied to the study of a completely different system composed of textile fibers in sea and fresh water. In this case the SAXS model taken into consideration was the one developed by Pedersen and Schurtenberger and further adapted to our system, allowing the investigation of the internal structure of microfibers with a particular focus on macrofibrils and microfibrils. By means of this approach, the behaviour of the textile fibers chosen for this analysis have been studied, and speculation about their mechanism of degradation as well as the contribution of UV radiation during this process have been obtained. On the whole, by means of all the models used and the systems analysed we can claim that the versatility and the contribution of the SAXS technique are crucial for the investigation of the most different scientific fields.

Bibliography

[1] Refractive index increment dn/dc values. URL https://www.materials-talks.com/refractive-index-increment-dndc-values/

[2] Ultimate guide to light scattering instruments for absolute macromolecular and nanoparticle characterization (2017)

[3] Amenitsch, H., Rappolt, M., Kriechbaum, M., Mio, H., Laggner, P., Bernstorff, S.: First performance assessment of the small-angle X-ray scattering beamline at ELETTRA. J. Sync. Rad. **5**, 506–508 (1998)

[4] Avio, C.G., Pittura, L., d'Errico, G., Abel, S., Amorello, S., Marino, G., Gorbi, S., Regoli, F.: Distribution and characterization of microplastic particles and textile microfibers in adriatic food webs: General insights for biomonitoring strategies. Environmental Pollution **258**, 113,766 (2020)

[5] Baden, N., Terazima, M.: Intermolecular interaction of myoglobin with water molecules along the ph denaturation curve. The Journal of Physical Chemistry B **110**, 15,548–15,555 (2006)

[6] Baldini, G., Beretta, S., Chirico, G., Franz, H., Maccioni, E., P. Mariani, F.S.: Salt induced association of β-lactoglobulin studied by salt light, x-ray scattering. Macromolecules **32**, 6128–6138 (1999)

[7] Barbosa, L.R.S., Spinozzi, F., Mariani, P., Itri, R.: Small-angle x-ray scattering applied to proteins in solution, pp. 49–72. Wiley (2013)

[8] Barrows, A., Cathey, S.E., Petersen, C.W.: Marine environment microfiber contamination: Global patterns and the diversity of microparticle origins. Environmental pollution **237**, 275–284 (2018)

[9] BARTIG, D., LEMKEMEIER, K., FRANK, J., LOTTSPEICH, F., KLINK, F.: The archaebacterial hypusine-containing protein: Structural features suggest common ancestry with eukaryotic translation initiation factor 5a. European journal of biochemistry **204**(2), 751–758 (1992)

[10] Bassani, F.: New functions for anarchaeal translation factor. Ph.D. thesis, Life and Environmental Sciences, Ancona (2017)

[11] Bassani, F., Romagnoli, A., Cacciamani, T., Amici, A., Benelli, D., Londei, P., Märtens, B., Bläsi, U., La Teana, A.: Modification of translation factor aif5a from sulfolobus solfataricus. Extremophiles **22**(5), 769–780 (2018)

[12] Benne, R., Brown-Luedi, M., Hershey, J.: Purification and characterization of protein synthesis initiation factors eif-1, eif-4c, eif-4d, and eif-5 from rabbit reticulocytes. Journal of Biological Chemistry **253**(9), 3070–3077 (1978)

[13] Berman, H., Henrick, K., Nakamura, H.: Announcing the worldwide Protein Data Bank. Nature Structural Biology **10**, 980–980 (2003)

[14] Blommel, P.G., Fox, B.G.: A combined approach to improving large-scale production of tobacco etch virus protease. Protein expression and purification **55**(1), 53–68 (2007)

[15] Browne, M.A., Crump, P., Niven, S.J., Teuten, E., Tonkin, A., Galloway, T., Thompson, R.: Accumulation of microplastic on shorelines woldwide: sources and sinks. Environmental science & technology **45**, 9175–9179 (2011)

[16] Campbell, I.D.: Biophysical Techniques. Oxford University Press (2012)

[17] Chen, M., Gai, Z., Okada, C., Ye, Y., Yu, J., Yao, M.: Flexible nad+ binding in deoxyhypusine synthase reflects the dynamic hypusine modification of translation factor if5a. International journal of molecular sciences **21**, 5509 (2020)

[18] Chen, Y.H., Ranganathan, P., Chen, C.W., Lee, Y.H., Rwei, S.P.: Effect of bis (2-aminoethyl) adipamide/adipic acid segment on polyamide 6: Crystallization kinetics study. Polymers **12**, 2073–4360 (2020)

[19] D'Agostino, M.: Structural and biochemical characterization of the aif5a-dhs complex. Ph.D. thesis, Life and Environmental Sciences, Ancona (2022)

[20] Defaye, A.B., Ledward, D.A.: Pressure-induced dimerization of metmyoglobin. Journal of Food Science **60**, 262–264 (1995)

[21] Dreyer, L.S., Nygaard, J., Malik, L., Hoeg-Jensen, T., Høiberg-Nielsen, R., Arleth, L.: Structural insight into the self-assembly of a pharmaceutically optimized insulin

analogue obtained by small-angle x-ray scattering. Molecular pharmaceutics **17**(8), 2809–2820 (2020)

[22] Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K., Kihara, H.: Evidence of an associative intermediate on the myoglobin refolding pathway. Biophys. J. **65**, 912–917 (1993)

[23] Falk, B.T., Liang, Y., McCoy, M.A.: Profiling insulin oligomeric states by 1h nmr spectroscopy for formulation development of ultra-rapid-acting insulin. Journal of Pharmaceutical Sciences **109**, 922–926 (2020)

[24] Feigin, L.A., Svergun, D.I.: Structure analysis by small-angle X-ray, neutron scattering. New York, Plenum Press (1987)

[25] Frauenfelder, H., Hartmann, H., Karplus, M., Kuntz Jr, I.D., Kuriyan, J., Parak, F., Petsko, G.A., Ringe, D., Tilton Jr, R.F., Connolly, M.L.: Thermal expansion of a protein. Biochemistry **26**, 254–261 (1987)

[26] Frühwirth, T., Fritz, G., Freiberger, N., Glatter, O.: Structure and order in lamellar phases determined by small-angle scattering. J. Appl. Cryst. **37**, 703–710 (2004)

[27] Ganim, Z., Jones, K.C., Tokmakoff, A.: Insulin dimer dissociation and unfolding revealed by amide i two-dimensional infrared spectroscopy. Phys. Chem. Chem. Phys. **12**, 3579–3588 (2010)

[28] Glatter, O., Kratky, O.: Small angle x-ray scattering. Academic Press, New Press (1982)

[29] Goldman, J., Carpenter, F.H.: Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. Biochemistry **13**, 4566–4574 (1974)

[30] Graewert, M.A., Franke, D., Jeffries, C.M., Blanchet, C.E., Ruskule, D., Kuhle, K., Flieger, A., Schäfer, B., Tartsch, B., Meijers, R., et al.: Automated pipeline for purification, biophysical and x-ray analysis of biomacromolecular solutions. Scientific reports **5**, 1–8 (2015)

[31] Habibi, Y., Lucia, L.A., Rojas, O.J.: Cellulose nanocrystals: Chemistry, self-assembly, and applications. Chemical Reviews **110**, 3479–3500 (2010)

[32] Haider, R., Sartori, B., Radeticchio, A., Wolf, M., Dal Zilio, S., Marmiroli, B., Amenitsch, H.: μDrop: a system for high-throughput small-angle X-ray scattering measurements of microlitre samples. Journal of Applied Crystallography **54**, 132–141 (2021)

[33] Hammersley, A.P.: *FIT2D*: a multi-purpose data reduction, analysis and visualization program. J. Appl. Cryst. **49**(2), 646–652 (2016)

[34] Hansen, J.P., McDonald, I.R.: Theory of Simple Liquids. Academic Press, London (1976)

[35] Hollóczki, O., Gehrke, S.: Nanoplastics can change the secondary structure of proteins. Scientific reports **9**, 1–7 (2019)

[36] Jeffrey, P.D., Coates, J.H.: An equilibrium ultracentrifuge study of the self-association of bovine insulin*. Biochemistry **5**, 489–498 (1966)

[37] Kamerzell, T.J., Esfandiary, R., Joshi, S.B., Middaugh, C.R., Volkin, D.B.: Protein–excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. Advanced drug delivery reviews **63**, 1118–1159 (2011)

[38] Kell, G.S.: Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. J. Chem. Eng. Data **20**, 97–105 (1975)

[39] Kourkoumelis, N., El-Gaoudy, H., Varella, E., Kovala-Demertzi, D.: Physicochemical characterization of thermally aged egyptian linen dyed with organic natural dyestuffs. Applied Physics A **112**, 469–478 (2013)

[40] Lin, Y.W., Wang, J.: Structure and function of heme proteins in non-native states: A mini-review. Journal of Inorganic Biochemistry **129**, 162–171 (2013)

[41] Manjunath, H., Zhang, H., Rehfeld, F., Han, J., Chang, T.C., Mendell, J.T.: Suppression of ribosomal pausing by eif5a is necessary to maintain the fidelity of start codon selection. Cell reports **29**(10), 3134–3146 (2019)

[42] Mathews, M.B., Hershey, J.W.: The translation factor eif5a and human cancer. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1849**, 836–844 (2015)

[43] Matsuoka, H., Tanaka, H., Hashimoto, T., Ise, N.: Elastic scattering from cubic lattice systems with paracrystalline distortion. Phys. Rev. B **36**, 1754–1765 (1987). DOI 10.1103/PhysRevB.36.1754

[44] Maurus, R., Overall, C.M., Bogumil, R., Luo, Y., Mauk, A., Smith, M., Brayer, G.D.: A myoglobin variant with a polar substitution in a conserved hydrophobic cluster in the heme binding pocket. BBA - Protein Structure and Molecular Enzymology **1341**, 1–13 (1997)

[45] Moretti, P.: Innovative methods to investigate intrinsically disordered proteins by x-ray and neutron scattering techniques. Ph.D. thesis, Life and Environmental Sciences, Ancona (2019)

[46] Moretti, P., Mariani, P., Ortore, M.G., Plotegher, N., Bubacco, L., Beltramini, M., Spinozzi, F.: Comprehensive structural and thermodynamic analysis of prefibrillar wt α-synuclein and its g51d, e46k, and a53t mutants by a combination of small-angle x-ray scattering and variational bayesian weighting. Journal of Chemical Information and Modeling **60**, 5265–5281 (2020)

[47] Murty, K.G.: Linear programming. Wiley Subscription Services, Inc., A Wiley Company (1983)

[48] Nagao, S., Osuka, H., Yamada, T., Uni, T., Shomura, Y., Imai, K., Higuchi, Y., Hirota, S.: Structural and oxygen binding properties of dimeric horse myoglobin. Dalton Trans. **41**, 11,378–11,385 (2012)

[49] Nettleton, E.J., Tito, P., Sunde, M., Bouchard, M., Dobson, C.M., Robinson, C.V.: Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. Biophysical Journal **79**, 1053–1065 (2000)

[50] Nishiyama, Y.: Structure and properties of the cellulose microfibril. Journal of Wood Science **55**, 241–249 (2009)

[51] O'Donoghue, S.I., Chang, X., Abseher, R., Nilges, M., Led, J.J.: Unraveling the symmetry ambiguity in a hexamer: Calculation of the r6 human insulin structure. Journal of Biomolecular NMR **16**, 93–108 (2000)

[52] Oehme, D.P., Doblin, M.S., Wagner, J., Bacic, A., Downton, M.T., Gidley, M.J.: Gaining insight into cell wall cellulose macrofibril organisation by simulating microfibril adsorption. Cellulose (6), 3501–3520 (2015)

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

[54] Ono, K., Ito, M., Hirota, S., Takada, S.: Dimer domain swapping versus monomer folding in apo-myoglobin studied by molecular simulations. Phys. Chem. Chem. Phys. **17**, 5006–5013 (2015)

[55] van den Oord, A.H.A., Wesdorp, J.J., van Dam, A.F., Verheij, J.A.: Occurrence and nature of equine and bovine myoglobin dimers. European Journal of Biochemistry **10**, 140–145 (1969)

[56] Ortore, M.G., Mariani, P., Carsughi, F., Cinelli, S., Onori, G., Teixeira, J., Spinozzi, F.: Preferential solvation of lysozyme in water/ethanol mixtures. J. Chem. Phys. **135**, 245,103–245,111 (2011)

[57] Ortore, M.G., Sinibaldi, R., Spinozzi, F., Carbini, A., Carsughi, F., Mariani, P.: Looking for the best experimental conditions to detail the protein solvation shell in

a binary aqueous solvent via Small Angle Scattering. Journal of Physics: Conference Series **177**, 012,007 (2009)

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

[59] Park, M.H.: The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5a (eif5a). Journal of biochemistry **139**(2), 161–169 (2006)

[60] Pedersen, J.S., Schurtenberger, P.: Scattering Functions of Semiflexible Polymers with and without Excluded Volume Effects. Macromolecules **29**, 7602–7612 (1996)

[61] Penttilä, P.A., Rautkari, L., Österberg, M., Schweins, R.: Small-angle scattering model for efficient characterization of wood nanostructure and moisture behaviour. Journal of Applied Crystallography **52**(2), 369–377 (2019). DOI 10.1107/S1600576719002012

[62] Peura, M., Müller, M., Vainio, U., Sarén, M.P., Saranpää, P., Serimaa, R.: X-ray microdiffraction reveals the orientation of cellulose microfibrils and the size of cellulose crystallites in single norway spruce tracheids. Trees **22**, 49–61 (2008)

[63] Piccinini, A., Lourenço, E.C., Ascenso, O.S., Ventura, M.R., Amenitsch, H., Moretti, P., Mariani, P., Ortore, M.G., Spinozzi, F.: Saxs reveals the stabilization effects of modified sugars on model proteins. Life **12**(1), 123 (2022)

[64] Prabhu, N.V., Sharp, K.A.: Heat capacity in proteins. Annual Review of Physical Chemistry **56**, 521–548 (2005)

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

[66] Quehenberger, J., Shen, L., Albers, S.V., Siebers, B., Spadiut, O.: Sulfolobus–a potential key organism in future biotechnology. Frontiers in microbiology **8**, 2474 (2017)

[67] Robertson, A.D., Murphy, K.P.: Protein structure and the energetics of protein stability. Chemical Reviews **97**, 1251–1268 (1997)

[68] Romagnoli, A.: Characterization of archaeal protein aif5a: a multifunctional translation factor. Ph.D. thesis, Life and Environmental Sciences, Ancona (2017)

[69] Romagnoli, A., Moretti, P., D'Agostino, M., Rexha, J., Perta, N., Piccinini, A., Di Marino, D., Spinozzi, F., Teana, A.L.: Structural–functional relationship of the ribonucleolytic activity of aif5a from sulfolobus solfataricus. Biomolecules **12**(10), 1432 (2022)

[70] Sane, R., Sinz, M.: Introduction of drug metabolism and overview of disease effect on drug metabolism, pp. 1–19. Academic Press (2017)

[71] Schellart, W.: Rheology and density of glucose syrup and honey: Determining their suitability for usage in analogue and fluid dynamic models of geological processes. Journal of Structural Geology **33**, 1079–1088 (2011)

[72] Schellman, J.A.: A simple model for solvation in mixed solvents. Applications to the stabilization and destabilization of macromolecular structures. Biohpys. Chem. **37**, 121–140 (1990)

[73] Schellman, J.A.: The thermodynamics of the solvent exchange. Biopolymers **34**, 1015–1026 (1994)

[74] Schellman, J.A.: Protein Stability in Mixed Solvents: A Balance of Contact Interaction and Excluded Volume. Biophys. J. **85**, 108–125 (2003)

[75] Schellman, J.A.: Destabilization and stabilization of proteins. Q. Rev. Biophys. **38**, 351–361 (2005)

[76] Schnablegger, H., Singh, Y.: The SAXS Guide. Getting acquainted with principles. AntonPaar GmbH, Austria (2017)

[77] Sinibaldi, R., Ortore, M.G., Spinozzi, F., Carsughi, F., Frielinghaus, H., Cinelli, S., Onori, G., Mariani, P.: Preferential hydration of lysozyme in water/glycerol mixtures: A small-angle neutron scattering study. J. Chem. Phys. **126**, 235,101–235,109 (2007)

[78] Sorci, M., Belfort, G.: Insulin oligomers: Detection, characterization and quantification using different analytical methods. In: V.N. Uversky, Y.L. Lyubchenko (eds.) Bio-nanoimaging, pp. 233–245. Academic Press, Boston (2014)

[79] Spinozzi, F., Beltramini, M.: QUAFIT: A Novel Method for the Quaternary Structure Determination from Small-Angle Scattering Data. Biophys. J. **103**, 511–521 (2012)

[80] Spinozzi, F., Carsughi, F., Mariani, P., Saturni, L., Bernstorff, S., Cinelli, S., Onori, G.: Met-myoglobin association in dilute solution during pressure-induced denaturation: an analysis at pH 4.5 by high-pressure small-angle X-ray scattering. J. Phys. Chem. B **111**, 3822–3830 (2007)

[81] Spinozzi, F., Ferrero, C., Ortore, M.G., Antolinos, A.D.M., Mariani, P.: 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)

[82] Spinozzi, F., Gazzillo, D., Giacometti, A., Mariani, P., Carsughi, F.: Interaction of proteins in solution from small angle scattering: a perturbative approach. Biophys. J. **82**, 2165–2175 (2002)

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

[84] Spinozzi, F., Ortore, M.G., Nava, G., Bomboi, F., Carducci, F., Amenitsch, H., Bellini, T., Sciortino, F., Mariani, P.: Gelling without structuring: A saxs study of the interactions among dna nanostars. Langmuir **36**, 10,387–10,396 (2020)

[85] Spinozzi, F., Paccamiccio, L., Mariani, P., Amaral, L.Q.: Melting Regime of the Anionic Phospholipid DMPG: New Lamellar Phase and Porous Bilayer Model. Langmuir **26**, 6484–6493 (2010)

[86] Sugiyama, J., Vuong, R., Chanzy, H.: Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. Macromolecules **24**, 4168–4175 (1991)

[87] Svergun, D., Richard, S., Koch, M.H.J., Sayers, Z., Kuprin, S., Zaccai, G.: Protein hydration in solution: experimental observation by X-ray, neutron scattering. Proc. Natl. Acad. Sci. USA **95**, 2267–2272 (1998)

[88] Svergun, D.I., Petoukhov, M.V., Koch, M.H.J.: Determination of domain structure of proteins from x-ray solution scattering. Biophys. J. **80**, 2946–2953 (2001)

[89] Uversky, V.N., Garriques, L.N., Millett, I.S., Frokjaer, S., Brange, J., Doniach, S., Fink, A.L.: Prediction of the association state of insulin using spectral parameters. Journal of Pharmaceutical Sciences **92**, 847–858 (2003)

[90] Waseda, Y., Matsubara, E., Shinoda, K.: X-ray diffraction crystallography: introduction, examples and solved problems. Springer (2011)

[91] Wertheim, M.S.: Exact solution of the percus-yevick integral equation for hard spheres. Phys. Rev. Lett. **10**, 321–323 (1963)

[92] Womersley, R.: Efficient spherical designs on the sphere s^2 with good geometric properties (2015). URL `https://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/sf.html`

[93] Womersley, R.S.: Efficient Spherical Designs with Good Geometric Properties, pp. 1243–1285. Springer International Publishing (2018)

[94] Yao, M., Ohsawa, A., Kikukawa, S., Tanaka, I., Kimura, M.: Crystal structure of hyperthermophilic archaeal initiation factor 5a: a homologue of eukaryotic initiation factor 5a (eif-5a). Journal of biochemistry **133**(1), 75–81 (2003)

[95] Zangi, R., Berne, B.J.: Temperature dependence of dimerization and dewetting of large-scale hydrophobes: A molecular dynamics study. The Journal of Physical Chemistry B **112**, 8634–8644 (2008)

[96] Zhou, C., Qi, W., Lewis, E.N., Carpenter, J.F.: Characterization of sizes of aggregates of insulin analogs and the conformations of the constituent protein molecules: A concomitant dynamic light scattering and raman spectroscopy study. Journal of Pharmaceutical Sciences **105**, 551–558 (2016)

[97] Zhu, K., Jia, H., Sun, Y., Dai, Y., Zhang, C., Guo, X., Wang, T., Zhu, L.: Long-term phototransformation of microplastics under simulated sunlight irradiation in aquatic environments: roles of reactive oxygen species. Water Research **173**, 115,564 (2020)