UNIVERSITÀ POLITECNICA DELLE MARCHE

Dipartimento di Scienze della Vita e dell'Ambiente

Corso di laurea Magistrale in Biologia Molecolare Applicata

curriculum Tecnologie Biologiche

Studi mediante diffusione dinamica di luce e diffusione a piccolo angolo dei raggi X della stabilità termica di proteine in presenza di zuccheri modificati

Studies through dynamic light scattering and small-angle X-ray scattering of protein thermal stability in the presence of modified sugars

Tesi di laurea Magistrale di:

Matteo Vercelli

Relatore: Chiar.mo

Prof. Francesco Spinozzi

Correlatore:

Prof. Yuri Gerelli

Sessione estiva

Anno Accademico 2021/2022

ABSTRACT

Le proteine in soluzioni acquose sono, in genere, poco stabili. Esse, infatti, sono sensibili a vari fattori, sia legati alla loro preparazione sia alle condizioni in cui vengono mantenute. La stabilità delle proteine dipende dal tipo di tampone, dal pH, dalla loro concentrazione e dallo stress meccanico al quale possono essere sottoposte. Per tali motivi, dal punto di vista biotecnologico, è importante mantenere le proteine in condizioni ottimali durante ogni processo, dalla produzione al trasporto, dallo stoccaggio fino al loro impiego. Il mantenimento delle condizioni ottimali comporta, in genere, difficoltà e costi elevati, soprattutto per mantenere, duranti i vari processi, le condizioni di bassa temperatura. Questo aspetto è di fondamentale importanza soprattutto per farmaci basati su molecole di natura proteica. Questi composti vanno mantenuti nelle appropriate condizioni per evitare la perdita della loro funzione e, in molti casi, anche il paziente deve adoperare delle accortezze per conservare l'integrità dei farmaci di natura proteica.

Allo scopo di rendere più semplice e meno costoso il mantenimento delle proteine, sono stati studiati diversi agenti da aggiungere in soluzione per stabilizzare le proteine. Questo studio di tesi è focalizzato sull'uso di quattro zuccheri modificati, prodotti dall'azienda portoghese ExtremoChem, i quali dovrebbero avere la capacità di stabilizzare le proteine, impedendo la loro

aggregazione. Le proteine usate in questo studio sono state l'insulina, la mioglobina ed ACE2, il recettore umano del virus SARS-CoV-2. Gli zuccheri modificati aggiunti alle soluzioni acquose delle proteine sono stati impiegati in diverse concentrazioni. I campioni sono stati successivamente studiati con due tecniche biofisiche: l'insulina e la mioglobina sono state analizzate tramite Dynamic Light Scattering (DLS), la proteina ACE2 è stata studita mediante Small Angle X-Ray Scattering (SAXS). Nel primo caso, si ricava la taglia delle proteine, nel secondo caso è possibile risalire sia alla loro dimensione che alla loro conformazione. Dei quattro zuccheri usati, due hanno caratteristiche ioniche, gli altri due non hanno cariche elettriche. Le misure di DLS sono state analizzate in termini di distribuzione volumica in funzione del raggio delle proteine. Da questa informazione, è possibile ricavare le popolazioni degli oligomeri formati dalle proteine. In particolare, per la mioglobina si sono determinate le frazioni di monomeri, dimeri ed strutture non ripiegate, mentre, per l'insulina, si sono determinate le frazioni di monomeri, dimeri, tetrameri ed esameri. Dai risultati ottenuti dai campioni di proteina con aggiunta dei quattro zuccheri modificati, è emerso che gli zuccheri non carichi sono quelli che maggiormente stabilizzano le proteine. Al contrario, gli altri due zuccheri con caratteristiche ioniche provocano aggregazione. Infatti, con i primi due zuccheri modificati, la mioglobina è presente per la maggior parte in forma di

monomero, e, per quanto riguarda la mioglobina, prevale la presenza di monomeri e dimeri. Diversamente, con gli zuccheri carichi, si nota la presenza di forme aggregate più grandi. Nel caso dell'insulina, la presenza della più alta concentrazione di zuccheri carichi prova precipitazione.

La proteina ACE2 è stata studiata in presenza di due dei quattro zuccheri, uno ionico e uno non carico. In primo luogo, dall'analisi grafica dei dati SAXS è emerso che lo zucchero non carico è quello con un effetto migliore sulla stabilizzazione della proteina, in accordo con quanto ottenuto dai dati DLS. Infatti, in presenza dello zucchero non carico, le curve SAXS sono simili a quelle della proteina nativa. Diversamente, nel caso dello zucchero carico, i dati SAXS mostrano che la proteina cambia conformazione e quindi è meno stabile rispetto al caso dello zucchero non carico. L'analisi dei dati SAXS mediante semplici modelli, mostra che ACE2 è costituita da un dominio globulare e un dominio non ripiegato. Il raggio di girazione del dominio compatto risulta essere maggiore del raggio della proteina nello stato monomerico. Ciò indica che ACE2 in soluzione potrebbe formare particelle costituite da più monomeri.

Index

Chapter I

INTRODUCTION

Proteins are not always stable in solution. They are sensible to different conditions, such as temperature increasing, pH variations, high concentration and other chemical-physical parameters. For these reasons, by a biotechnological point of view it is important to maintain the optimal conditions in all the steps from proteins' production to transport and storage. However, this control involves high costs and difficulties since, , for example, proteins need to be kept at low temperature. On these premises, it is of fundamental importance to stabilize proteins against denaturation or aggregation. For this purpose, chemical compounds, typically small molecules, which are dissolved in water solution together with proteins and are able to stabilize them against denaturation are deeply investigated. This agents can help to preserve and protect the structural integrity, and hence the functionality, of proteins.

Some proteins play the role of therapeutic agents. Thanks to their effects, they are applied in the treatment of several diseases. For example, insulin's injections are fundamental for people suffering from diabetes. But, like many other peptide hormone based drugs, insulin is temperature sensitive thus it suffers storage troubles and needs continuous refreshing [1]. As a matter of

fact, temperature stability of proteins is one of the main limitations in their use as potential drugs. Proteins can also be degraded by other factors, such as high saline concentrations, mechanical stresses etc, which determine the loss of the three-dimensional structure. Hence, stabilizing conditions, in general, aim to preserve the proper protein conformation that, if lost, causes the loss of protein activity. Proteins in solution are characterized by a melting temperature (T_m), above which they rapidly lost the tertiary structure. The melting temperature represents a changing point associated with the free energy. Above T_m, non-covalent bonds, like hydrogen bond and salt bridges, are broken and there is a conversion from folded to unfolded state and this is related with a free energy change which is negative. The result is that the denatured state is thermodynamically favoured. Temperature increase above T_m leads a series of conformational changes such as denaturation, covalent and noncovalent aggregation and oxidation.

Proteins are formed by a specific sequence of amino acids, each with its characteristics, due to the distinct side chains. In particular, amino acids are distinguished on the basis of their hydrophilic or hydrophobic nature. Generally, the hydrophobic regions of a protein in its folded and native state are folded inward to avoid the contact with solvent. When protein unfolds, the hydrophobic residues result exposed to the solvent and the only way to counter

this process is to bind hydrophobic areas of other proteins, leading to aggregation [2].

Proteins in solution are surrounded by a hydration layer build by water molecules that are more ordered and locked than the ones forming the bulk solvent. This layer has a very important role in proteins stability, since it controls the structure and therefore the protein functionality. The solvation water may be disturbed by the presence of a cosolvent; this perturbance could be both negative and positive. In some cases, cosolvents destabilize the water shell and can determine a variation of the protein structure from its native state. However, in other cases proteins are stabilized by cosolvents, which improve the resistance to degradation [3]. These aspects suggest that a promising avenue to stabilize the protein structure is to add in solution cosolvent molecules able to preserve the hydration layer. Compounds used as stabilizers are mostly small molecules, able to interact with the thin layer of water around the protein. These stabilizers can be identified in sugars, salts, amino acids and polymers. Amino acids are active on proteins stabilization in different ways based on its their characteristics. For example histidine is an antioxidant and protect proteins from oxidative damage thanks to its free radical cleavage. Ions that form salt dissolution take action accumulating on protein surface and prevent the unfolding [4].

Among these stabilizers, sugars are the most studied as well as used. Studies have revealed that some sugars, dissolved in a protein solution, improve protein stability preventing aggregation [5]. Sugars mainly act on hydration shell and do not directly enter in contact with protein. In fact, they are mainly accumulated in the bulk water so that their presence gives rise to a reinforcement of the water shell encompassing protein, by producing an overall stabilizing effect on protein structure [6]. This happens because of the hydrogen bond, formed between protein and water in the hydration shell, cannot be disturbed because the other water molecules in solution are captured by sugars [7].

In this thesis work we tested the effect of four modified sugars on proteins stability. These have been synthesized by ExtremoChem, a Portuguese company that develops and produces various stabilizers acting on bio-macromolecules and they could result really useful on maintaining protein structure against high temperatures, pH variations, mechanical stresses and high concentration. These modified sugars (coded as EC-101, EC-202, EC-212, EC-311, EC-312) contain mannose, glucose or galactose with different substituents. Substituents have various characteristics that influence their stabilization effect [8]. In particular, EC-212, EC-311 and EC-312 are

uncharged, whereas EC-101 and EC-202 are salts, hence they form ionic species in solutions.

This work is based on a previous study [9] which investigated the role of the modified sugars, described above, on the stabilization of insulin and myoglobin as model proteins. In that study, authors analysed samples with small-angle X-Ray scattering (SAXS) technique as a function of temperature and sugars concentration, with the purpose to quantify the distribution of proteins in different states. In a similar manner, we here investigated insulin and myoglobin in presence of sugars and by varying temperature by means of dynamic light scattering (DLS) technique. In addition we studied the protein ACE2, which has a great interest in the scientific community because of its role on SARS-CoV-2 infection. This was also analysed in relation with sugars but experiments were performed by a synchrotron SAXS technique.

1.1 Myoglobin

In our body, oxygen needs to be transported and stocked. This is implemented by some proteins that are able to bind molecules of oxygen. Oxygen arrives in the lung through breath and then it takes contact with blood in the alveoli. Here there is an exchange: erythrocytes release carbon dioxide

and they bind oxygen which is transported in the organs. Haemoglobin is the protein contained in the erythrocytes that allows oxygen transport. Then, in the tissues, oxygen is transferred on myoglobin, with the task of storing it. Myoglobin is more present in skeletal muscles and heart because almost all the oxygen is consumed by those two tissues [10]. Both proteins bind oxygen thanks to heme group, a factor consisting of an iron-porphyrin complex inserted into the protein structure. Haemoglobin, with a molecular weight of 67 kDa, is made up of four chains and it contains four heme groups. Myoglobin is a smaller protein, present as monomer with a molecular weight of around 17 kDa and containing one heme group [11].

Myoglobin is a single chain formed by 153 amino acid residues and the sequence is (*Figure 1.1*) :

```
>1WLA_1|Chain A|MYOGLOBIN|Equus caballus
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKT
EAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIK
YLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG
```

Figure 1.1 : In figure is reported myoglobin sequence in FASTA format. 1WLA is the PDB query and Equus caballus is the organism to which analysed protein belongs.

From a structural point of view, myoglobin has eight α-helixes connected by some loops (*Figure 1.2*). The exposed surface is hydrophilic, whereas the inward surfaces contain hydrophobic residues.

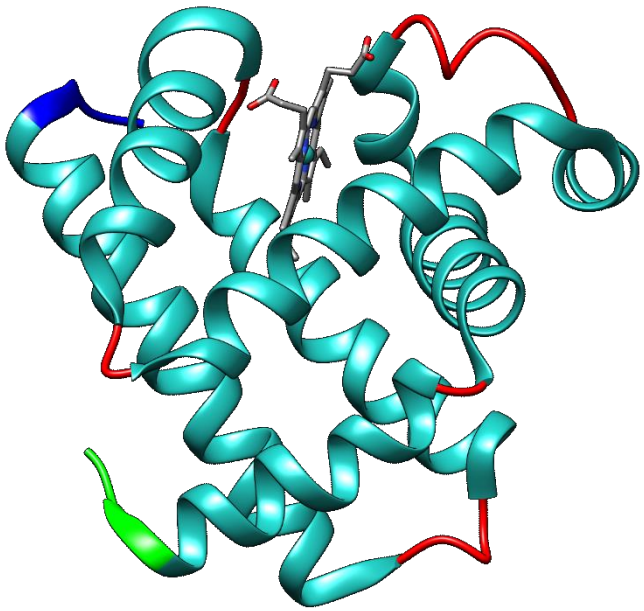


Figure 1.2 : Myoglobin represented with ribbon to highlight the helixes. In red are reported loops, N-terminal is in green and C-terminal is the blue one.

The stability of this protein depend on intrinsic forces given by hydrogen bond, van der Waals forces and hydrophobic interactions. Moreover, the molecule is stabilized by the interactions with the heme group [12], which is located in a hydrophobic pocket. This group can be considered a disk of high electron density formed by porphyrin rings, with an iron atom in the centre (*Figure 1.3*). The heme group is attached to the protein through the interactions of the iron

atom with the nitrogen atom of the side chain of a histidine residue (*Figure 1.4*) [13] [14].

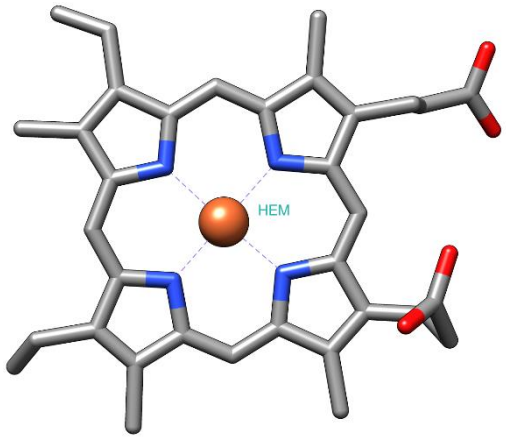


Figure 1.3 : Heme group. Four nitrogen atoms by porphyrin rings stabilize the iron atom.

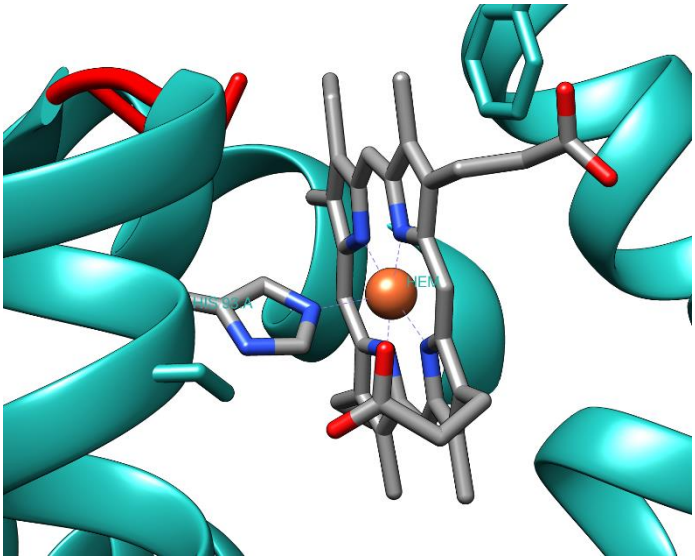


Figure 1.4 : Heme group is bound at the protein through a nitrogen atom in the histidine 93.

The role of myoglobin is the oxygen transport from the erythrocyte to the mitochondria in cells like cardiomyocytes and skeletal myofibers. In fact it

was seen that myoglobin loses its oxygen in proportion to exercise intensity. During exercise, cells need more oxygen to produce more ATP and this is given by myoglobin to mitochondria.

Myoglobin is also a functional reservoir for oxygen and it helps the mitochondrial oxidative phosphorylation. This protein may also function as a nitric oxide scavenger and has a positive effects on the elimination of oxide free radical [15].

1.2 Insulin

Different districts of human body communicate thanks to a molecular system which consists in hormones transported through the blood. Insulin is one of the most important hormone and it transports information that indicate the sugars concentration in blood. Insulin is made in the pancreas and is released after meals with a high sugars level. Insulin spreads into the entire organism and binds its receptors on cells of liver, muscle and fat cells. The insulin's task is to induce these organs to eliminate glucose from the blood, so it induces organs to adsorb glucose and to store it in glycogen form or fat.

In particular, insulin is produced by β-cells in pancreas, which are electrically excitable in response to glucose. When glucose arrives on the cells

there is a release of Ca^{2+} ions with paracrine and autocrine signaling, so it acts among the β-cells also. Under this stimulus, there is a insulin secretory response. During its travel, the first organ that insulin meets is liver. Therefore, liver is exposed to high insulin concentrations and it always acts in the regulation of insulin levels in the other organs. In particular these levels are directed at muscles and adipose tissues where insulin increases the glucose uptake and leads the storage of glucose in glycogen form [16].

The biosynthesis of insulin follows two steps: indeed it contains two chains, A and B, in mature form but it is produced in a single chain polypeptide called proinsulin. Proinsulin has the B-chain in the N-terminal and the A-chain in C-terminal, connected by a segment in the middle called C-peptide. A-chain and B-chain are bound thanks to three disulphide-bonds, two of them between the two chains and the third internal into the A-chain. Subsequently, a proteolytic process occurs, which eliminates the C-peptide and both the A-chain remain, with 21 amino acids, and the B-chain, with 30 amino acids, are formed (*Figure 1.5*) [17].

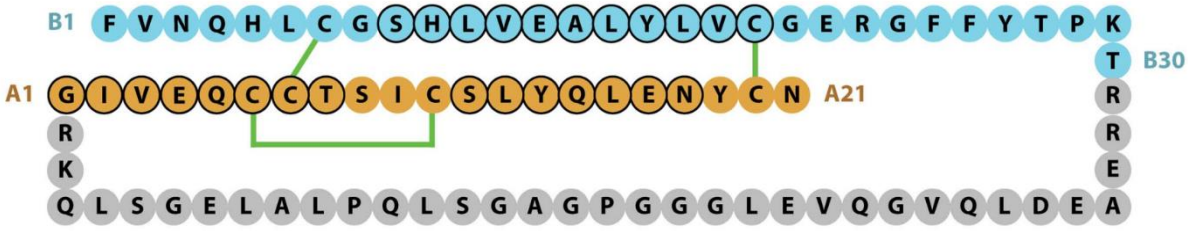


Figure 1.5 : Orange, light blue and grey residues representing the A-chain, the B-chain and the C-peptide of insulin, respectively, this latter being eliminated by proteolysis. The green lines indicate disulphide-bonds. The amino acids circled in black are the ones which form an α-helix.

After the proteolytic process, the protein assumes its structure defined by two α-helixes in A-chain and one α-helix and one β-sheet in B-chain, in addition to loops that keeps together all the secondary structures (*Figure 1.6*) [18].

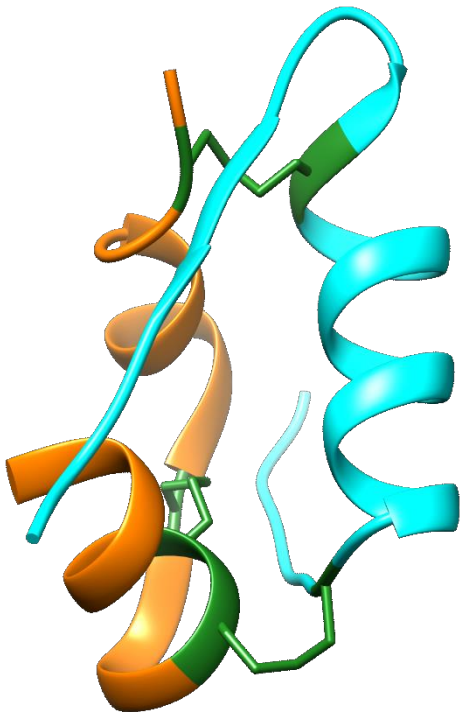


Figure 1.6 : Insulin structure, in orange is represented the A-chain and the B-chain in light blue. In green are the disulphide-bind inter and intrachain.

Starting from the monomer, insulin can produce superstructure as dimer, tetramer and hexamer. The formation of these compounds is lead by zinc atoms which is bound to the histidine 10 of the B-chain and the interaction between zinc atoms of different monomer allows the formation of larger structure (*Figure 1.7*) [19].

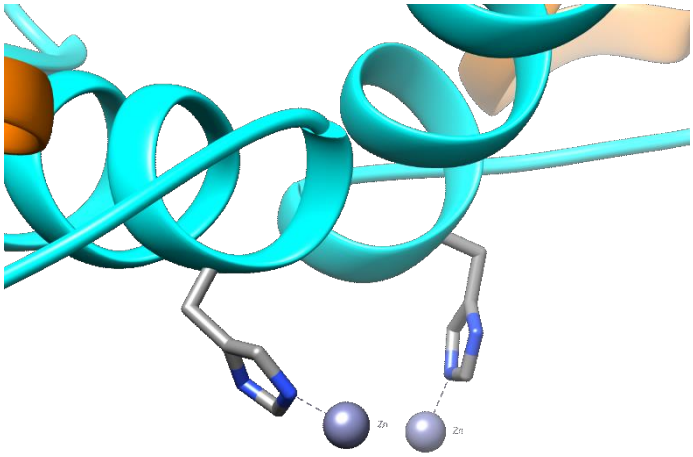


Figure 1.7 : Histidine in position 10 in B-chain attached a zinc atom.

The zinc atoms of two monomer interact with each other and allow the formation of the dimer. The zinc atom can take contact with more than one molecule, so it coordinates different monomers and dimers and, in this way it can provoke the formation of larger structure. The interaction of two dimers origins the tetramer and when three dimers interact the hexamer is formed. Two zinc atom are able to attach three dimers and to lead the formation of an hexamer (*Figure 1.8*) [20].

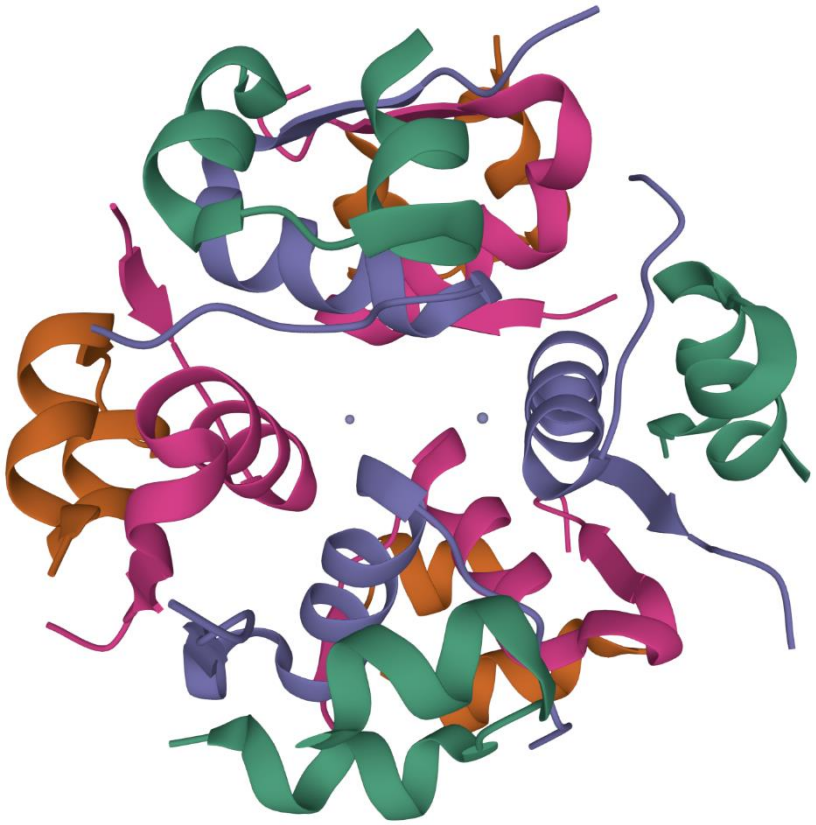


Figure 1.8 : Insulin hexamer. Each zinc atom interacts with three monomer.

In human organism insulin, is really important. In fact, when insulin production system is impaired, for example for a pancreatic damage, there are continuously high levels of glucose in blood and this leads the arise of diabetes mellitus. This disease is really dangerous for people totally deficient in insulin production, such as children. Diabetes causes a lack of glucose uptake in cells, so the high glucose levels in blood lead to dehydration, because the body tries to eliminate the excess of sugars with urine. Moreover dangerous changes in the pH value of blood can happen, because the body has no longer available

sugars, hence it uses other acid molecules as energy sources. Sugar molecules, spreading in the whole body through the blood, can interact with proteins, by provoking the loss of their function, and accumulate on the cells, by harmful effect. People suffering from diabetes may be treated by manually supplementing the insulin. These people need an insulin source for use in the treatment. Insulin is produced by the extraction of pig or cow insulin which are very similar to the human one. Thanks to the improvements in biotechnological field, a recombinant human insulin from bacteria can be produced, which does not differ from human insulin [21].

1.3 ACE2

From 2020 in the entire world there is one of the great pandemic of our era. This is caused by the virus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). SARS-CoV-2 gives a large spectra of manifestations related to patient's conditions, as previous diseases and the age. Sometimes infection is asymptomatic. In other cases, it manifests with an easy flu and, in the most severe situations, leads to death of the patient caused by interstitial pneumonia. During the last two years COVID-19 caused millions of deaths due to the poor epidemiological control in some countries.

SARS-CoV-2 belongs to the family of the *Coronaviridae*, which also includes other well-known viruses responsible for other epidemies as MERS and SARS and even it includes the common flu viruses.

Phylogenetic analysis on different coronavirus genomes reveals that the most similar genome, hence the nearest phylogenetically related, to the SARS-CoV-2 genome it is the virus that infects bats, called RaTG13. The gene sequence which leads the production of the spike protein is similar at 93% in the two viruses. For this reason, the scientific community retains that the COVID-19 pandemic was originated by one or more spill overs, which enabled the virus to pass through different species [22].

SARS-CoV-2 is an RNA (+) virus and codify for all the necessary proteins. In particular, there are four structural proteins, which are the transmembrane protein, the envelope protein, the one of nucleocapsid and the spike protein [23]. Among these, the spike protein is really important and the most studied. because it permits the binding with the human receptor ACE2 (Angiotensin-Converting Enzyme 2). The spike protein has three different sections with different functions. The section called S1 is the most protruding part and it contains the RBD domain (Receptor Binding Domain) that takes contact with the receptor ACE2. The section S2 carries out its role in the fusion

of virus with the host cell. The third part has a transmembrane domain, which connects the spike protein with the virus's surface.

ACE2 in human body regulates the cardiovascular homeostasis. It is located on the surface of the cells and it is mostly represented in organs such as lung, heart, kidney and intestine [24]. In particular, ACE2 belongs to the enzymatic class of hydrolases and it leads the conversion from Angiotensin II to Angiotensin - (1 - 7). The metabolic process begins from angiotensinogen that is a protein produced by liver. Angiotensinogen is converted in angiotensin I. thanks to the renin which is an hydrolase produced by kidneys. This hormone stimulates vasoconstriction and then it is transformed in angiotensin II by ACE enzyme. Thus ACE2 converts angiotensin II in angiotensin – (1-7), which balances the effects of angiotensin II [25].

ACE2 is a transmembrane glycoprotein formed by 805 amino acids. It has a molecular weight of 76 kDa and the mature protein consist in many subunits, being a hetero pentamer. In the *Figure 1.9,* bottom, it is possible to identify three domains: TMD is the transmembrane domain that anchors the protein to the membrane; PD is the extracellular catalytic site and the neck connect both sides [26].

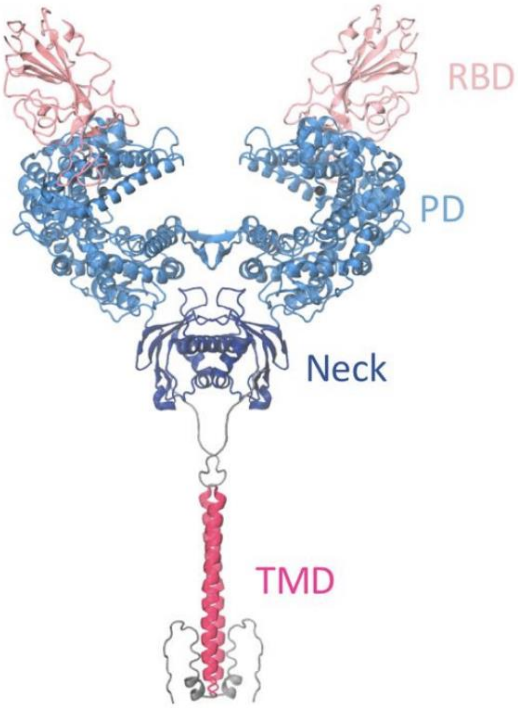


Figure 1.9 : Structure of ACE2 with the three domains PD (light blue), TMD (red) and Neck (blue). Also it is represented the binding with RBD domain of spike protein (pink).

In this Figure the domain of spike protein, which binds the receptor, is also reported. From the study of the complex spike – ACE2, in particular between RBD and PD, it has been observed that RBD contains five antiparallel β-sheets and, between two of them, there is a zone with α-helixes and loops called RBM. This RBM zone is responsible for the interaction between spike protein and the N-terminal domain of ACE2. Spike protein binds ACE2 in a lobed pocket and then some enzymes, which cause conformational changes that lead the fusion between the virus and the cell membrane, are involved [22].

In our experiments, we analysed a part of the domain PD of ACE2, in particular the soluble region of 35 kDa, which binds directly the spike protein (*Figure 1.10*) [27].

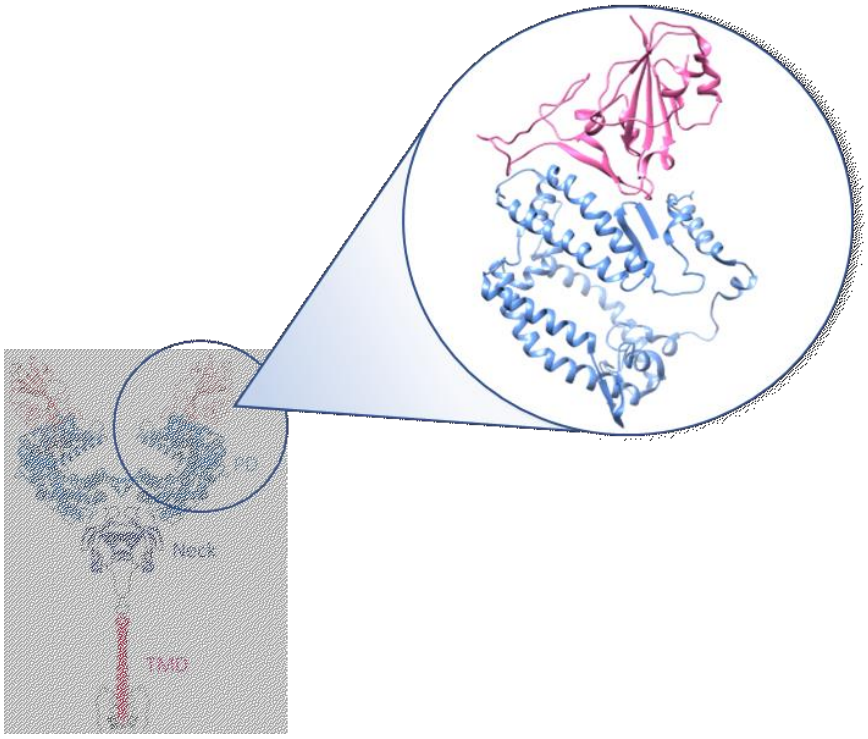


Figure 1.10 : This image represents the used part of PD (light blue) in our analysis and its contact with RBD domain of spike protein (pink).

The amino-acid sequence of ACE2 is reported below with the specific sequence for the soluble part of domain PD highlighted (*Figure 1.11*) :

>sp|Q9BYF1|ACE2_HUMAN Angiotensin-converting enzyme 2 OS=Homo sapiens
MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQ
NMNNAGDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTIL
NTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLY
EEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHL
HAYVRAKLMNAYPSYISPIGCLPAHLLGDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQ
AWDAQRIFKEAEKFFVSVGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILM
CTKVTMDDFLTAHHEMGHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKHLKS
IGLLSPDFQEDNETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKKWWEM
KREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYTRTLYQFQFQEALCQAAKHEGPLH
KCDISNSTEAGQKLFNMLRLGKSEPWTLALENVVGAKNMNVRPLLNYFEPLFTWLKDQNK
NSFVGWSTDWSPYADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMRQYFLKVKN
QMILFGEEDVRVANLKPRISFNFFVTAPKNVSDIIPRTEVEKAIRMSRSRINDAFRLNDN
SLEFLGIQPTLGPPNQPPVSIWLIVFGVVMGVIVVGIVILIFTGIRDRKKKNKARSGENP
YASIDISKGENNPGFQNTDDVQTSF

Figure 1.11: The sequence of protein region which was studied with SAXS technique is highlighted in green. Considering all the sequence, the first 17 amino acids codify for a signal peptide and the last 65 amino acids codify for the other domain such as the neck and the transmembrane domain.

Chapter II

AIM OF THE STUDY

In the previous chapter we discussed about protein denaturation and the importance to find some agents, to add at the protein preparation, able to stabilize it against denaturation or aggregation. This aspect is relevant for example for drugs conservation. From drugs preparation to its transport and storage until its use by ordinary people, it is necessary to maintain the cold chain. This is not easy and it is also expensive.

A common drug. such as insulin, used by everyone who suffers diabetes, requires continuous cooling and this is really complicate because patients need to carry with them insulin everywhere with the risk to ruin the drug because it is transported in cool bags that maintain low temperatures just for a few hours. This is not a big deal for the world areas with cold or moderate temperatures in which, even if insulin is transported in a cool bag, the low temperature is maintained for quite a long time. On the contrary, it is really difficult to preserve insulin at low temperature in tropical countries, such as in Africa or in some regions of South America.

These aspects encourage researchers to search and develop molecular agents which have the task to preserve drugs. For this reason we investigated

the effect of five different sugars on two model proteins and one protein of global interesting by means of different techniques. Insulin and myoglobin were analysed with DLS because, with the aim to determine protein aggregation and quantify the volume fraction of different protein states. ACE2 was analysed with SAXS, with and without stabilizing sugars. At the best of our knowledge, this was the first SAXS investigation of the soluble part of ACE2.

The overall aim of this study is to understand if the sugars, produced by ExtremoChem, are effective on protein stabilization and to classify their effects. This analysis was made to find an agent with a positive effect on proteins stabilization which is applicable to protein preparation and hence that can help the protein maintenance.

Chapter III

MATERIALS AND METHODS

3.1 Buffers

Myoglobin and Insulin samples were prepared using the same 10 mM phosphate buffer, respectively at pH 5 and pH 3, prepared following the protocol described in Piccinini et al. (2022). In this way, the same conditions used in previous SAXS experiments were also employed for the DLS experiments performed in this study.

After preparation, each buffer was filtered through a 0.2 μm Cellulose Acetate filter, in order to eliminate dust particles and other possible impurities.

3.1.1 Myoglobin's buffer

Myoglobin was dissolved in phosphate buffer 10 mM at pH 5. The buffer was prepared using :

- Phosphoric acid 85%
- NaOH 1 M (enough to reach pH 5)
- Deionized water

3.1.2 Insulin's buffer

Insulin was dissolved in phosphate buffer 10 mM at pH 3. The buffer was prepared using:

- Phosphoric acid 85%
- NaOH 1M (enough to reaches pH 3)
- Deionized water

3.1.3 ACE2's buffer

For ACE2 samples, a 10 mM PBS buffer at pH 7.4 was used. This buffer was prepared using :

- NaCl 137 mM
- KCl 2.7 mM
- Na_2HPO_4 10 mM
- KH_2PO_4 1.8 mM

3.1.4 Sugars' buffer

Sugars were dissolved in either phosphate buffer at pH 5, used for Myoglobin samples, or phosphate buffer at pH 3, used for Insulin samples.

Experiments on ACE2 was performed at a later date, so we just had sugars dissolved in buffer at both pH, without possibility to use new dried

sugars to dissolve in the same buffer of ACE2. For this reason, for ACE2 samples, sugars dissolved in phosphate buffer at pH 5 were used, since they did not lower too much the pH of the solution.

3.2 Stocks' preparation

After preparation, stocks of MB and IN were filtered with 0.2 μm filter to remove contaminant particulate that would strongly affect the DLS measurements.

3.2.1 Myoglobin's stock

Myoglobin from equine heart was purchased from Sigma-Aldrich. It was dissolved in the relative buffer at a concentration of 7.5 g/L.

3.2.2 Insulin's stock

Insulin from bovine pancreas was acquired from Sigma-Aldrich. This protein was dissolved in buffer at pH 3 in a concentration of 2.4 g/L.

3.2.3 Stock of ACE2

ACE2 was produced in the NY-MaSBiC laboratory as recombinant protein expressed in bacterial system and purified from culture of these bacteria [27]. A stock of ACE2 solution at 4 g/L was used.

3.2.4 Sugars' stocks

Each sugar (EC-101, EC-202, EC-212, EC-311, EC-312) was dissolved in both phosphate buffers at a concentration of 1.5 M.

After filtration, we needed to check if sugars were still present in solution. For this reason, we performed a UV spectrophotometric analysis of EC-202 before and after filtration. The measured absorption spectra are shown in *Figure 3.1*.

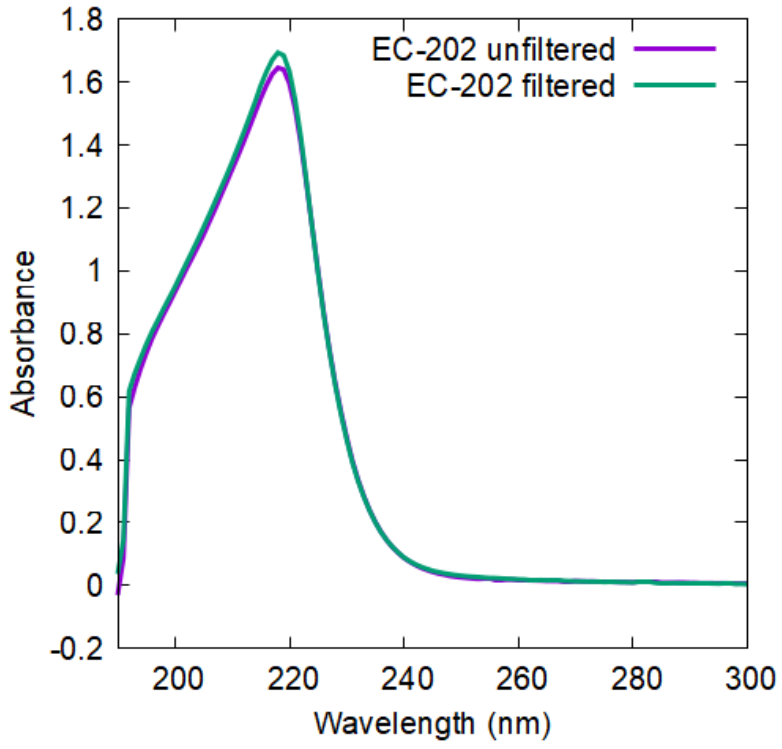


Figure 3.1: UV spectra of EC-202 stock solution before and after the filtration.

The two spectra are very similar, with a maximum absorption peck at 220 nm, confirming that the filtration process does not remove sugars by the solution.

3.3 Sample's preparation for DLS

Only four of the five available sugars were used, since the EC-311 powder resulted to be altered. For each protein concentration, samples were prepared at 0.05 M, 0.1 M and 0.25 M sugar concentration by mixing proper amount of protein and sugar stocks. To remove protein aggregates, the protein stock was filtered before sugars addition.

Myoglobin samples were prepared at 2 g/L and 4 g/L, whereas Insulin samples were prepared only at 2 g/L. The list of investigated samples is hereafter shown:

- MB 2 g/L + EC-101 0.05 M
- MB 2 g/L + EC-101 0.1 M
- MB 2 g/L + EC-101 0.25 M
- MB 2 g/L + EC-202 0.05 M
- MB 2 g/L + EC-202 0.1 M
- MB 2 g/L + EC-202 0.25 M
- MB 2 g/L + EC-212 0.05 M

- MB 2 g/L + EC-212 0.1 M
- MB 2 g/L + EC-212 0.25 M
- MB 2 g/L + EC-312 0.05 M
- MB 2 g/L + EC-312 0.1 M
- MB 2 g/L + EC-312 0.25 M
- MB 4 g/L + EC-101 0.05 M
- MB 4 g/L + EC-101 0.1 M
- MB 4 g/L + EC-101 0.25 M
- MB 4 g/L + EC-202 0.05 M
- MB 4 g/L + EC-202 0.1 M
- MB 4 g/L + EC-202 0.25 M
- MB 4 g/L + EC-212 0.05 M
- MB 4 g/L + EC-212 0.1 M
- MB 4 g/L + EC-212 0.25 M
- MB 4 g/L + EC-312 0.05 M
- MB 4 g/L + EC-312 0.1 M
- MB 4 g/L + EC-312 0.25 M
- IN 2 g/L + EC-101 0.05 M
- IN 2 g/L + EC-101 0.1 M
- IN 2 g/L + EC-101 0.25 M

- IN 2 g/L + EC-202 0.05 M
- IN 2 g/L + EC-202 0.1 M
- IN 2 g/L + EC-202 0.25 M
- IN 2 g/L + EC-212 0.05 M
- IN 2 g/L + EC-212 0.1 M
- IN 2 g/L + EC-212 0.25 M
- IN 2 g/L + EC-312 0.05 M
- IN 2 g/L + EC-312 0.1 M
- IN 2 g/L + EC-312 0.25 M

3.4 Sample's preparation for SAXS

SAXS experiments on ACE2 samples were performed at 1.5 g/L protein concentration in the presence and in the absence of either EC-202 or EC-312 sugars at 0.05 M, 0.1 M and 0.25 M. Investigated samples are listed below:

- ACE2 1.5 g/L + EC-202 0.05 M
- ACE2 1.5 g/L + EC-202 0.1 M
- ACE2 1.5 g/L + EC-202 0.25 M
- ACE2 1.5 g/L + EC-312 0.05 M
- ACE2 1.5 g/L + EC-312 0.1 M

- ACE2 1.5 g/L + EC-312 0.25 M

3.5 Dynamic Light Scattering (DLS)

3.5.1 Introduction to DLS

Light scattering is a technique used mainly for the characterization of particle size in diluted colloidal systems. It can be applied on diluted solutions containing particles with dimensions ranging from few nanometers up to several micrometers. In particular, the main result of a Dynamic Light Scattering experiment is the intensity-weighted size distribution of particles in solution.

Particles can give two different types of scattering, related to their dimension. According to Rayleigh scattering, when the particle size is lower than $1/10^{th}$ of the wavelength λ of the incident light, the light is uniformly scattered in any direction and maintains the same energy of the incident light. This process determines an isotropic scattering with an intensity proportional to the 6^{th} power of particles' radius (*Figure 3.2*). In the case of particles with size greater than $\lambda/10$, the scattering light is mainly scattered in the forward direction and the intensity depends on the scattering angle by a complex function that shows many maxima and minima, according to the particle shape.

The last two situations are referred to as Mie scattering processes (*Figure 3.2; Figure 3.3*) [28] [29].

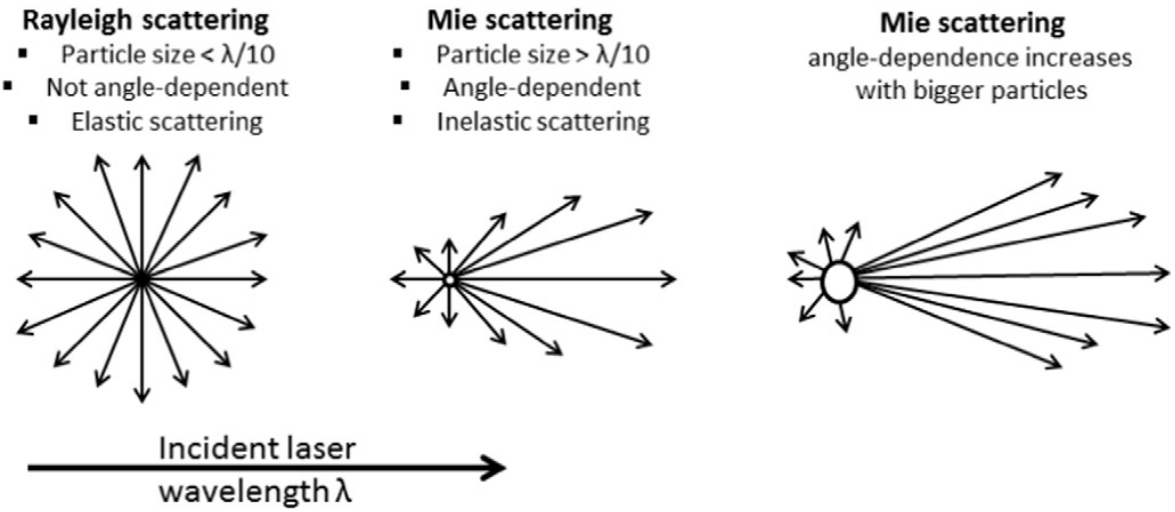


Figure 3.2 : Differences between Rayleigh and Mie scattering.

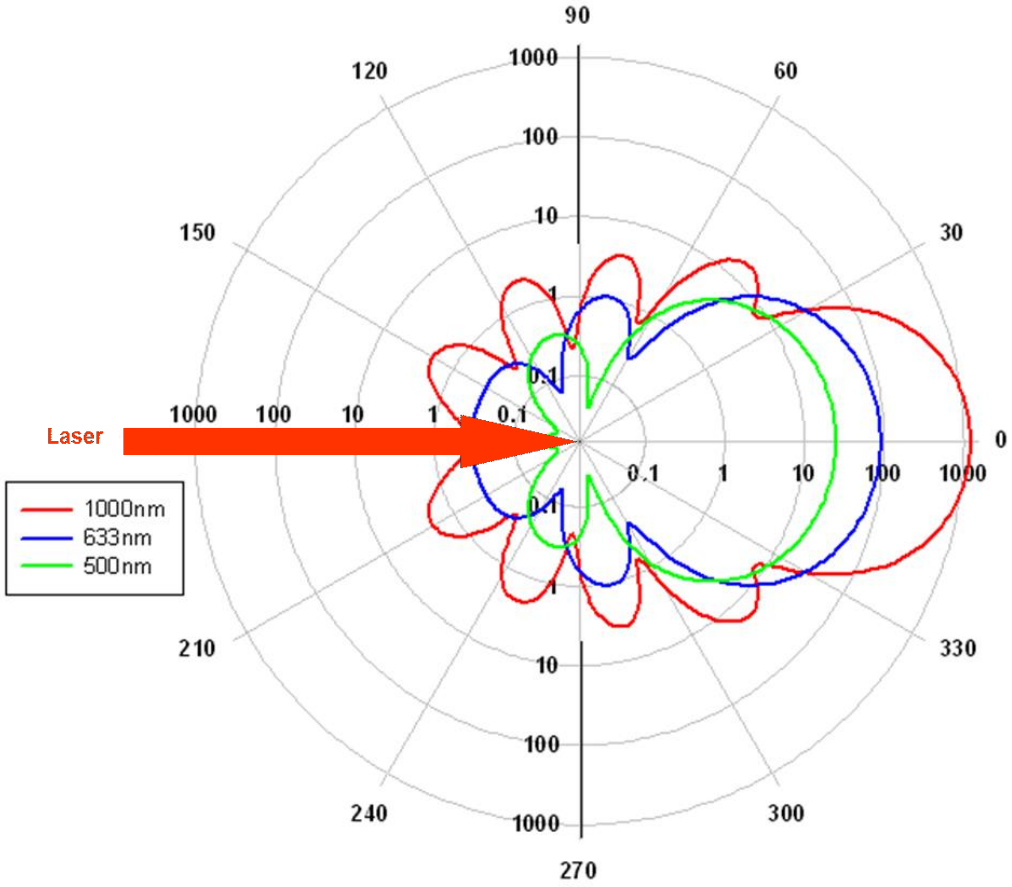


Figure 3.3 : Polar plot representing the Mie scattering of three particles with different size. The bigger is the size, the bigger is the signal. The majority of the scattering occurs in the forward direction. © 2019 Malvern Panalytical.

If particles do not move within the volume probed by the instrument, the scattered light, at a given angle, will be characterized by a constant intensity value. On the contrary, if particles are moving, marked fluctuations of the intensity of the scattered light will appear. This is the common case for particles in solution as they experience the Brownian motion, corresponding to random movements of particles due to stochastic collisions among them as well as with solvent molecules. Intensity fluctuations are measured during the experiment

by means of an intensity correlation function that is characterized by a decay profile dictated by the particles' translational diffusion coefficient (D). The latter defines the velocity of Brownian motion, and it relates to the particles size according to the Stokes-Einstein equation (*Eq. 1*).

$$d_H = \frac{k_B T}{3\pi\eta D}$$

d_H is the particle hydrodynamic diameter, k_B is Boltzmann's constant, T and η are the absolute temperature and the viscosity of the solution, respectively, and D is the particles' diffusion coefficient [28] [29]. From (*Eq. 1*) one can derive that the diffusion rate of small particle is higher than that of larger ones.

The hydrodynamic diameter is basically considered the diameter of a hard sphere that diffuses at the same rate of the particles being measured. The hydrodynamic diameter might not correspond to the true size of a particle and differences depend on several factors such as ionic strength, surface structure and shape. For example, a particle in a low ionic strength solution is surrounded by an extended electrical layer (*Figure 3.4A*). This situation reduces the diffusion speed and leads to a larger apparent hydrodynamic diameter. On the other hand, a particle in a high ionic strength solution has a compact electrical layer, so it appears as a smaller particle (*Figure 3.4B*).

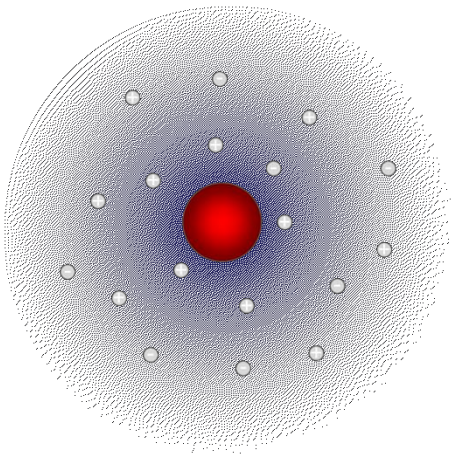


Figure 3.4A : Molecule with low ionic strength. Its occupies a larger surface. © 2019 Malvern Panalytical.

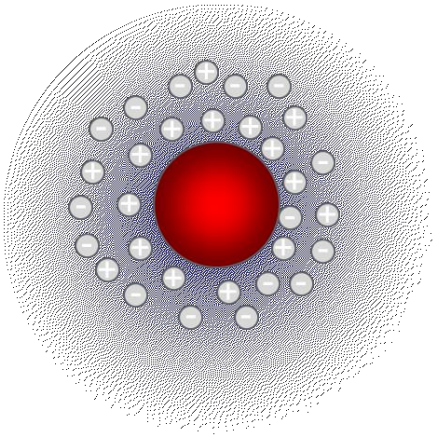


Figure 3.4B : Molecule with high ionic strength. This molecule appear as a smaller one. © 2019 Malvern Panalytical.

3.5.2 Malvern's instrument: Zetasizer pro

For our DLS analysis the Zetasizer pro (Malvern Panalytical) instrument has been used. The instrument is equipped by a He-Ne laser with $\lambda = 633$ nm. The incident light is directed toward the sample, placed in a cuvette, and the scattering light can be collected by a photo-detector at either 173° (back scattering angle), or 13° (forward scattering angle) measured in respect to the direction defined by the incident laser beam (*Figure 3.5*).

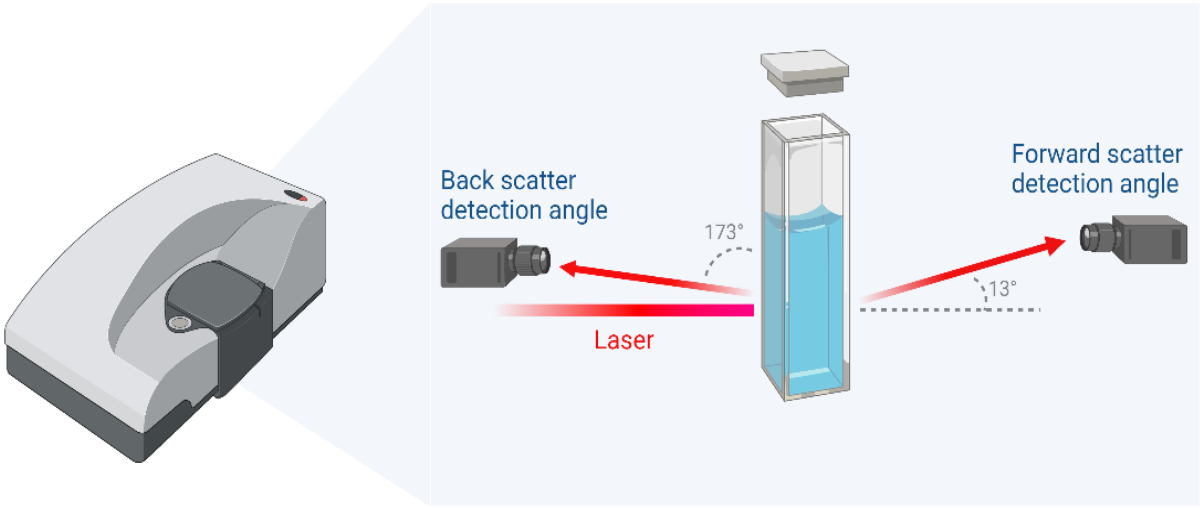


Figure 3.5 : Schematic representation of overall functioning of DLS with detectors' positions.

The scattering light intensity detected as a function of time, $I(t)$, is shown on a graph called count rate which reports the number of photons arriving at the detector per second (*Figure 3.6*).

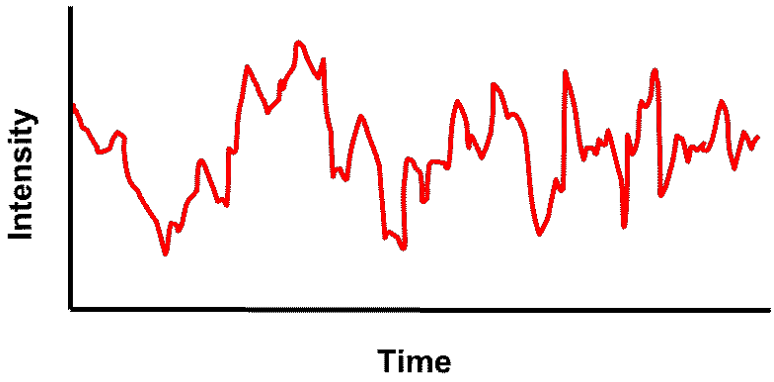


Figure 3.6 : Representation of a general count rate. © 2019 Malvern Panalytical .

From several measurements of $I(t)$, the average time correlation function $G(\tau)$ is calculated according to

$$G(\tau) = \langle \frac{I(t_0) * I(t_0 + \tau)}{I(t_\infty)^2} \rangle \qquad (2)$$

where τ is defined as the delay time. Suppose to translate, in time, the count rate plot of a certain amount; this is the delay time τ and it is possible to measure the correlation between the original count rate and the translated one. At τ = 0 no differences are expected (C(τ=0)=1); when τ value increases, the overlap between the original signal and the translated one will decrease therefore. The larger is the difference, the lower is the correlation function calculated with Eq. 2. Such a behavior is represented in *Figure 3.7*.

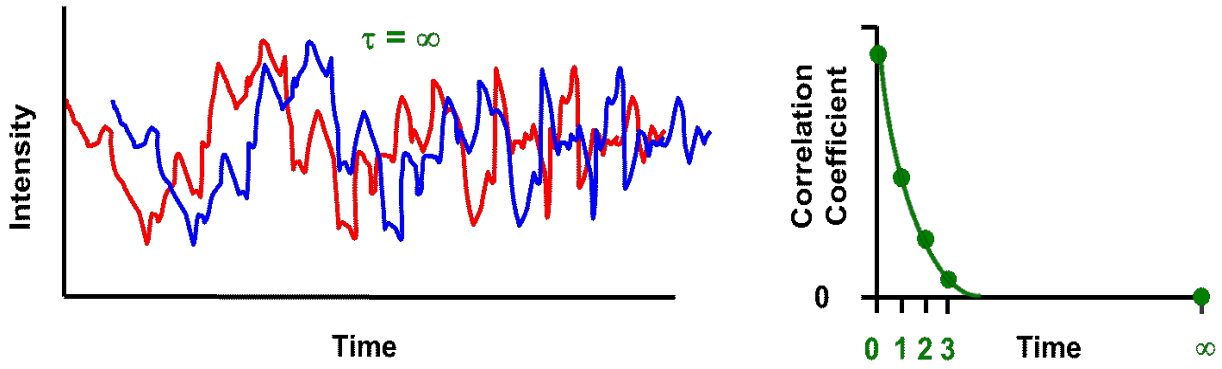


Figure 3.7 : Graphic explanation of the use of the delay time to concert the count rate into a correlation function. © 2019 Malvern Panalytical.

As the correlation function is expected to show an exponential decay profile, it is most commonly displayed in a lin-log scale as in the case of the plot calculated by the Malvern software during the measurements (*Figure 3.8*).

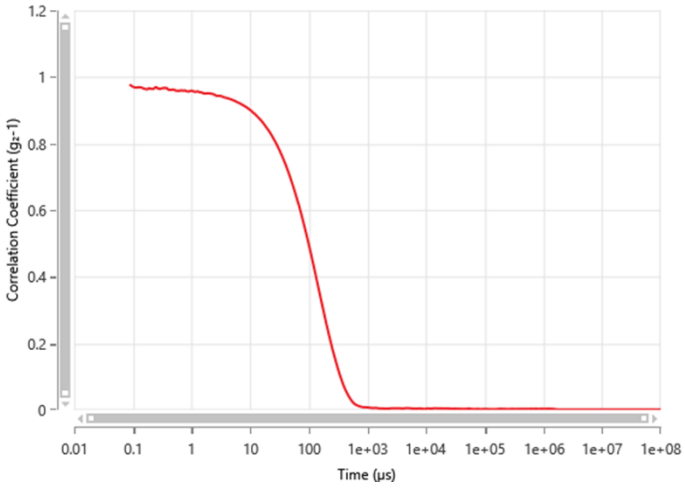


Figure 3.8 : Correlogram example from software ZS XPLORER. © 2019 Malvern Panalytical.

A typical correlogram should extrapolate to unity for $\tau=0$ while one or multiple exponential decays should be present if particles in solution are characterized by diffusion constants and size compatible with those accessed by the instrument. The time when the correlogram starts to decay is related to the mean size of the particles. The slope of the line during decay depend on the

polydispersity of particles' sizes: the faster is the decay, the lower is the polydispersity.

It is possible to convert the correlation function into a graph that reports the size distribution related to the scattered light intensity, using different methods, including numerical and analytical ones. The formula (*Eq. 3*) :

$$G_2(\tau) - 1 = A\left(\sum_i e^{-q^2 D_i \tau}\right)^2 \qquad (3)$$

can be used to extract the diffusion constants D_i of all the particles in the samples. Here A plays the role of an amplitude, and q is the modulus of the scattering vector defined as :

$$q = \frac{4\pi n}{\lambda_0} \sin\frac{\theta}{2} \qquad (4)$$

In this equation, n is the dispersant refractive index, λ the laser wavelength and θ the scattering angle. Note that the ensemble of D_i values corresponds to an intensity weighted diffusion constant distribution that can be further converted into an intensity weighted (hydrodynamic) size distribution according to Equation 1.

An alternative method for the analysis of the correlation function is the computation of the corresponding intensity weighted size distribution using

CONTIN-based methods. Similar methods are included in the Malvern software but details about them are not given by the developer given their proprietary nature.

For the distribution analysis, the instrument uses three different models. It is possible to select the appropriate algorithm to perform the more suitable analysis. They have different characteristics related to the sample. We choose the algorithms called L-curve analysis. It is considered the more apt for protein samples, because it reports narrow peaks and is able to automatically select the right distribution.

It is important to highlight that the intensity weighted size distribution, $I(R)$, is related to the intensity of light scattered by the particles. For example, using Rayleigh approximation, the intensity is proportional to R^6. Hence, if $N(R)$ is the number weighted size distribution, we have the simple relation $I(R) \propto R^6 N(R)$. In many circumstances, it is useful to calculate the volume weighted size distribution, which is related to $N(R)$ by $V(R) \propto R^3 N(R)$, the volume of the particle being proportional to the third power of its radius. It is thus straightforward to calculate the distributions $N(R)$ and $V(R)$ from the distribution $I(R)$ obtained by DLS data analysis, just by dividing for R^6 or R^3, respectively. To note, the volume *fraction* distribution $\varphi(R)$ can be obtained by dividing $V(R)$ by its integral $\varphi(R) = V(R) / \int_0^\infty V(R)\, dR$. This function can

be used to derive the fraction of the volume occupied by all the particles due the particles with radius comprised between R_1 and R_2: $p(R_1, R_2) = \int_{R_1}^{R_2} \varphi(R)\, dR$.

The determination of $\varphi(R)$ from DLS data is particularly useful in the case of protein aggregation. The reasons are shown hereafter. Let us suppose that a monomeric protein can form different oligomers. Let us call x_i the fraction of all the monomers that are aggregated in form of i-oligomers, where $i = 1$, stands for monomers, $i = 2$ for dimers and so on and so forth. Clearly $\sum_{i=1} x_i = 1$. If in the solution there are N_m monomers, independently on their aggregation state (i.e. $N_m = cN_A/M_1$, where c is the protein w/v concentration, M_1 is the monomer molecular weight and N_A is Avogadro's numbers), the number of proteins that remain in the monomeric state will be $x_1 N_m$, the number of dimers will be $x_2 N_m/2$ and, in general, the number of i-oligomers will be $x_i N_m/i$. Accordingly, the volume occupied by all the i-oligomers is the volume of one oligomer, iV_1 (V_1 is monomer's volume), times the number of i-oligomers, that is $(x_i N_m/i)(iV_1) = x_i V_1 N_m$. Hence the volume fraction due to the i-oligomers is the ratio between $x_i V_1 N_m$ and the sum $\sum_{j=1} x_j V_1 N_m$, which reads $x_i / \sum_{j=1} x_j$. So, it has been demonstrated that, for protein aggregation, the volume fraction corresponds to the fraction of all the monomers distributed

in oligomers. As a consequence, the integral $p(R_1, R_2)$, introduced above, also represents the fraction of the monomeric protein molecules that are forming oligomers with radius comprised between R_1 and R_2.

It is also worth to notice that in the ideal case of a system constituted by identical spheres, the intensity weighted size distribution $I(R)$, derived from DLS data analysis, and the related volume fraction distribution $\varphi(R)$ show a peak with a finite width. Notice that if the system is formed by identical particles with a non-spherical geometry, the peak shape can be asymmetric. Hence, in the case of protein distributed in many oligomers, the function $\varphi(R)$ will be characterized by a superposition of peaks (not necessarily symmetric), each one referring to a oligomeric species. And, on the basis of the discussion presented above, the area above each peak does represents the fraction of monomers that are constituting that oligomer.

These considerations suggest that a proper way to analyze the $\varphi(R)$ profiles is to fit them by a sum of peaks, for example Gaussian peaks. Hence we have the fitted the $\varphi(R)$ distribution according to the following relation,

$$\varphi(R) = \sum_{i=1}^{N_G} w_i e^{-(R-R_i)^2/(2\sigma_i^2)} \qquad (5)$$

where R_i is the peak center position, σ_i the peak standard deviation and w_i the relative weight. Notice that one oligomer could be represented by more than

one Gaussian peak, due the intrinsic asymmetry of protein particles. Hence the total number of Gaussians, N_G, would not necessarily corresponds to the number of oligomers.

In details, we can assign a group of subsequent Gaussian peaks, indexed, for example, from i_1 to i_2, to a certain oligomeric species, indexed by k. The fraction of monomers forming the k-oligomers will be given by

$$x_k = \sqrt{2\pi} \sum_{i=i_1}^{i_2} w_i \, \sigma_i \qquad (6)$$

where we have exploited the analytical expression of the area above a Gaussian function.

We have developed scripts under the Gnuplot software to calculate $\varphi(R)$ and, above all, to fit this data by *Eq. 5*.

3.5.3 Experiment's set up

Our experiments aim to describe the low-resolution structural behavior of proteins in water solution in the absence and in the presence of modified sugars and as a function of temperature. We planned to perform DLS measurements in from 20° to 70° C using heating and cooling temperature scans. Cooling scans were used to determine if any aggregation process that might occur was reversible. More in detail, for each sample composition,

measurements were subsequently performed at 20°, 30°, 40°, 50°, 60°, 70°, 60°, 40°, and 20° C.

For each temperature three repetitions of the measurement were performed in order to reach a more accurate data statistics as well as to check for sample stability. Also, when temperature was changed, samples were left 120 s to equilibrate before the beginning of the subsequent DLS measurement. Measurements were performed in back-scattering geometry with low-volume quartz cuvettes (ZEN2112), characterized by a small aperture with bottom restriction (*Figure 3.10*).

Figure 3.10 : Thanks to the small cell, it can use very low volume of sample. In backscatter, the minimum sample volume is 12 µl. © 2019 Malvern Panalytical.

As briefly stated before, it is very important to clean the cuvette and to operate with dust-free buffers. For this reason, the cuvette was washed with MilliQ water filtered through a 200 nm PVDF filter (for hydrophilic solutions), shacked and then the water was discarded. This procedure was repeated for three times. After this process, checks were regularly performed measuring filtered water (or buffer) in the clean cuvette. At room temperature, the Rayleigh scattering of clean water is expected to be almost isotropic with a value of 35 kcps (kilo counts per second) and the relative correlogram should not show any signal (*Figure 3.11*). On our Malvern instrument, a slight increase in the correlation function of pure water is always present and is considered to be caused by intrinsic instrumental limitations.

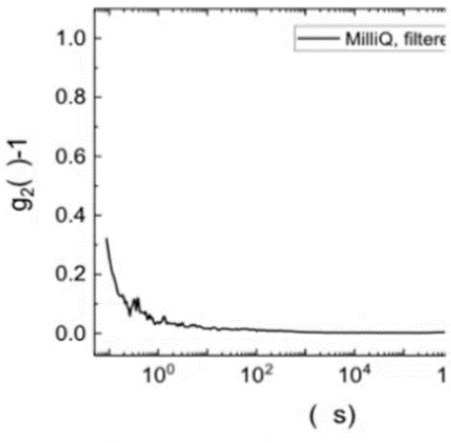


Figure 3.11 : Typical MilliQ correlogram does not give signal. This indicates there is nothing in the cell.

When cuvette was cleaned, we proceeded with the sample measurements inserting a volume of 70 μL of previously filtered sample in the clean and dry cell using a long tip to arrive at the cell's deep.

3.6 Small Angle X-Ray Scattering (SAXS)

3.6.1 Introduction

Small-angle X-Ray Scattering (SAXS) is an technique able to investigate the structure of macromolecules or assemblies of small molecules as well as their aggregation properties in solution. In particular, SAXS is considered one of the best biophysical technique to detect the structure of proteins in solution. With SAXS it is possible to analyze samples in almost any chemical-physical condition, in contrast with crystallography that requires that protein is crystallized. On the other hand, crystallography is the main technique for the determination of the protein structure at atomic resolution. Crystallography succeed to determine the exact atom's position in space, but also it has disadvantages. First of all, it requires crystallization of the molecules in a very organized structure, but not all proteins crystallize. Also, proteins blocked in the crystal are not able to interact with other molecules. For this reasons, SAXS

is an important approach for the analysis of proteins in solution, in a determinate temperature and interacting with other agents. SAXS has a low resolution but permits to obtain information about the shape and the interactions in the sample.

3.6.2 X-Rays production

There are two points of view to understand the light's nature. In a classic concept, a monochromatic light beam it is an electromagnetic wave with a certain wavelength λ. Such waves are formed by an electric field and an magnetic field that oscillate perpendicular to each other (*Figure 3.12*). In contrast, according to quantum point of view, light has a corpuscular nature, being represented by photons, which are particles moving, in vacuum, at the speed of light c and having energy only depending on the frequency $\nu = c/\lambda$ according to Plank's law $E = h\nu$ (h is Plank's constant). In particular, X-Ray are electromagnetic waves with wavelength in a range of 0.1 and 100 Å and also they are photons with energy between 0.1 and 100 keV.

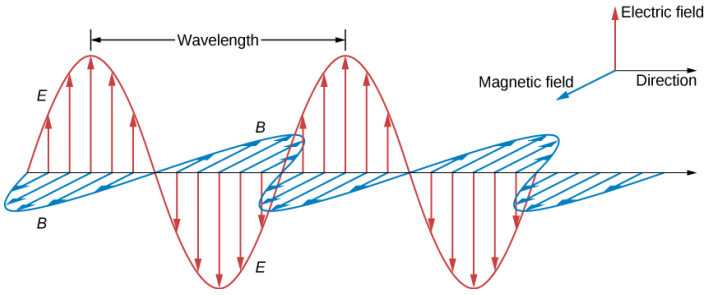


Figure 3.12 : Representation of an electromagnetic field.

X-Rays can be produced at very high brilliance in a synchrotron facility. Synchrotrons accelerate electrons close to the speed of light in a circular trajectory (the synchrotron's ring) in which there are some magnets. Magnets are able to swing the electrons and, when they change direction, they emit electromagnetic waves in a wide range of frequencies, including the ones of X-Rays. In a synchrotron there is also an electron gun, which generates the electrons and insert them in the so-called booster synchrotron, a smaller ring used for pre-accelerate the electrons before their injection in the storage ring. Along the circle there are also different magnets with different function (*Figure 3.13*) [30].

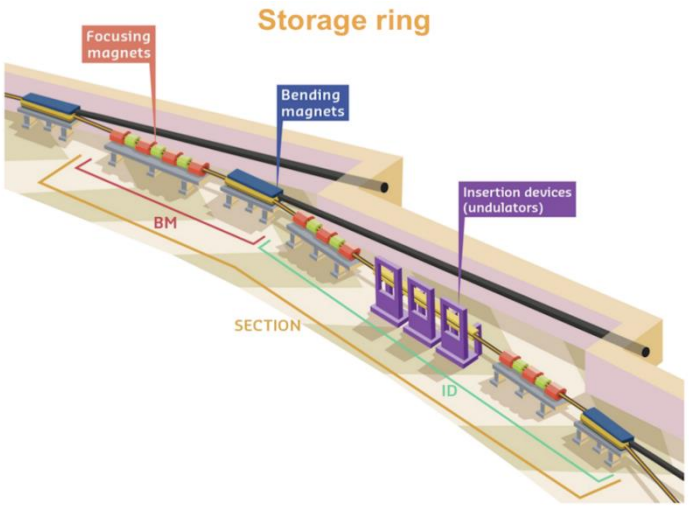


Figure 3.13 : Schematic representation of the magnets in the storage ring.

There are multiple focusing magnets in the circle which guarantee that electrons remain in their right orbit. Insertion devices, also called undulators, serve to change the electrons' straight line in an undulation trajectory. The consequence of this process is the emission of X-Rays (*Figure 3.14*) [30].

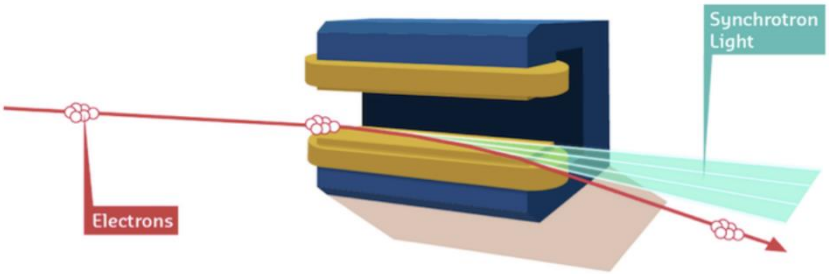


Figure 3.14 : Representation of a bending magnet. It produces a sort of spray containing a large spectrum of wavelength.

Finally, the bending magnets use a different technique to produce X-Rays, which are less focused than ones produced with insertion devices. In this case, electrons pass through the magnet and they are deflected from their straight line. This method produces a synchrotron light which covers a wide spectrum of light, from microwaves to X-Rays (*Figure 3.15*) [30].

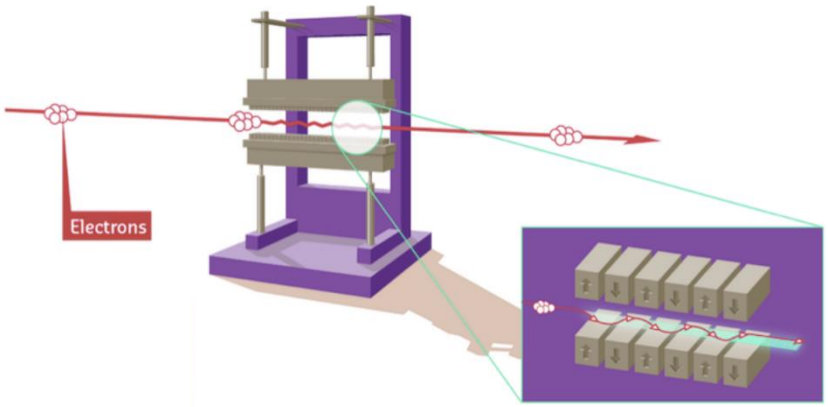


Figure 3.15 : When electrons reach insertion device run into a series of magnets with alternate orientation, this lead the undulation.

Due to the production of many different wavelengths, it is necessary to select no only the ones in the X-Rays range, but also a specific one, according to the experimental needs. This is possible using a monochromator, a device that contains a crystal. When an incident light arrives at the crystal is deflected with a certain angle that depends on the incident angle (*Figure 3.16*).

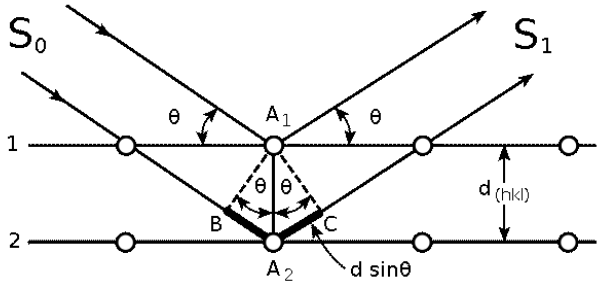


Figure 3.16 : Structure and function of a monochromator.

The incident beams S_0 hit crystal atoms in different layers, but the angle θ is the same. The resulting rays S_1 could give a constructive interference when the difference in optical path is equal to 2dsinθ, then there will be an increase of the signal, otherwise it gives a destructive interference. This process is ruled by Bragg's law (*Eq. 7*).

$$n\lambda = 2d \sin\theta \qquad (7)$$

The X-Rays produced are directed toward beamlines that encircles the storage ring. Beamlines are equipped with a set of tools. First of all, X-Rays enter in the optics cabin provided with mirrors and crystals used to focus the beam and select the right wavelength. Then X-Rays are directed in the experimental cabin. In this area analysis take place, so there is the instrument where are located the samples and the detectors to catch X-Rays collided with sample. At least, there is the control cabin. Here users can use software to control the beam, custom experiment, set any conditions and capture data.

Our SAXS experiment was performed at ESRF, the European Synchrotron, in Grenoble (*Figure 3.17*). In particular we used the BM29 beamline, dedicated to the study of proteins and macromolecules in solution (*Figure 3.18*) [31].

Figure 3.17 : ESRF, The European Synchrotron.

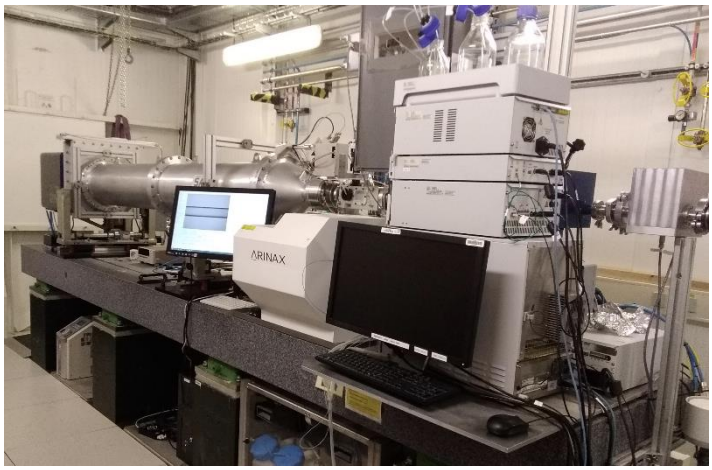


Figure 3.18 : Experimental cabin of beamline BM29.

3.6.3 SAXS theory

When a beam of X-Rays hits the sample, the electrons of the atoms forming molecules are excited. The vibration of electrons produce spherical waves (*Figure 3.19*), because electron vibrates in each direction. But in the sample there are many electrons and each of them produces spherical waves which create an interference with deflection of the incident beam. Signal is captured by the detector located at a distance of a few meters and it records the intensity of scattered radiation.

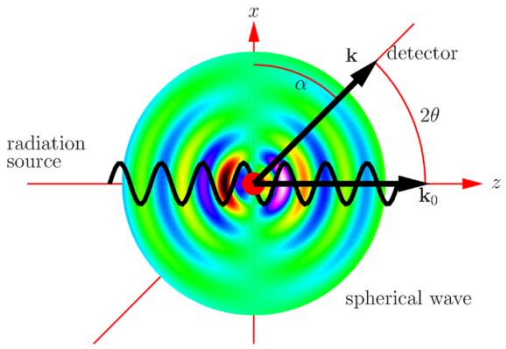


Figure 3.19 : Pattern of the spherical waves for a system with one electron.

Considering a system of two electrons hit by X-Rays at the same time. Distance between electron is defined by a position vector r. As can be seen in

the *Figure 3.20*, the scattered radiation consist in two parallel waves but with different optical path.

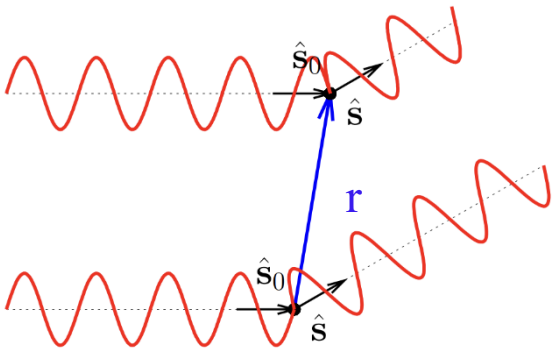


Figure 3.20 : Scattering of two electrons.

The unit vector in the incident beam direction is indicated by $\hat{s}_0$ and the one corresponding to the scattered beam is $\hat{s}$. The angle between incident and scattered direction is 2θ. The generated waves have the same wavelength λ, but difference in the optical path could create a phase different δ, giving constructive or destructive interference and is calculated with *Eq. 8* :

$$\delta = \frac{\vec{r}}{\lambda} \cdot (\hat{s} - \hat{s}_0) \qquad (8)$$

It is important to define the scattering vector $\vec{Q}$ that depends on unit vector difference (*Eq. 9*),

(*9*)

$$\vec{Q} = \frac{2\pi}{\lambda}(\hat{s} - \hat{s}_0)$$

The modulus of $\vec{Q}$ is obtained from *Eq. 9* and is calculated with *Eq. 10* :

$$Q = \frac{4\pi}{\lambda}\sin\theta \qquad (10)$$

SAXS intensities are recorded by the detector for different values of the scattering angle 2θ (which is typically very low, around some tenth of degrees) and, after a radial average that sum all the counts of the pixels that are at the same distance from the primary beam position into the detector, are reported as a function of the modulus Q.

Considering a monodisperse system (i.e. a system formed by a single type of particles), and assuming that the electron density ρ_p of the particles is mostly constant (such as in the case of proteins) and different from the one of the solvent, ρ_0, and that the particles are randomly oriented and very diluted in solution, the SAXS intensity (more precisely defined as the macroscopic differential scattering cross section) can be expressed by the following equation (*Eq. 11*)

$$\frac{d\Sigma}{d\Omega}(Q) = n_p(\Delta\rho)^2 V_p^2 P(Q) S_M(Q) \qquad (11)$$

In this equation there are several constants. They are: n_p the number of particle per unit volume (number density, related to the molar concentration C by $n_p = CN_A$, N_A being Avogadro's number); $\Delta\rho = \rho_p - \rho_0$, known as the electron density contrast (or just the contrast); V_p, the particle volume. Hence the SAXS signal is proportional to the particle concentration and to both the squared contrast and the squared particle volume. The dependency on Q is contained in the form factor $P(Q)$ and the "measured" structure factor $S_M(Q)$. This latter depends on the correlation among the particles: if the solution is diluted (with particle volume fraction in the order of 10^{-3}) $S_M(Q) \approx 1$, whereas *(12)* presence of protein attraction or repulsion, i.e. at moderate of high protein concentration, $S_M(Q)$ can be lower or higher than 1, respectively. The form factor is defined by the following expression (*Eq. 12*):

$$P(Q) = \langle \left| F\left(\vec{Q}\right) \right|^2 \rangle_{\alpha_Q, \beta_Q}$$

where the symbol $\langle ... \rangle_{\alpha_Q, \beta_Q}$ stands for the average over the polar angles of the vector $\vec{Q}$ and $F\left(\vec{Q}\right)$ represents the normalized amplitude of the wave scattered by the particle, which is the Fourier transform calculated only within the volume of the particle, according to (*Eq. 13*).

$$F_i\left(\vec{Q}\right) = \frac{1}{f_i} \int_{V_i} d^3\vec{r}\, \delta\rho_i(\vec{r}) e^{i\vec{Q}\cdot\vec{r}} \qquad (13)$$

where $\vec{r}$ stands for a position vector within the particle. Notice that, despite $F(\vec{Q})$ is a function in the field of complex numbers and depends on a vector, the function $P(Q)$, being defined by the modulus square of $F(\vec{Q})$, is a real and positive function. Moreover, the orientational average renders $P(Q)$ a function that only depends on the modulus Q that can be easily represented in a bidimensional plot. On the basis of its definition, it is evident that $P(Q)$ describes the shape as well as the size of the particle. An example is the plot shown in *Figure 3.21*, which reports the form factors of particles with different geometrical shapes.

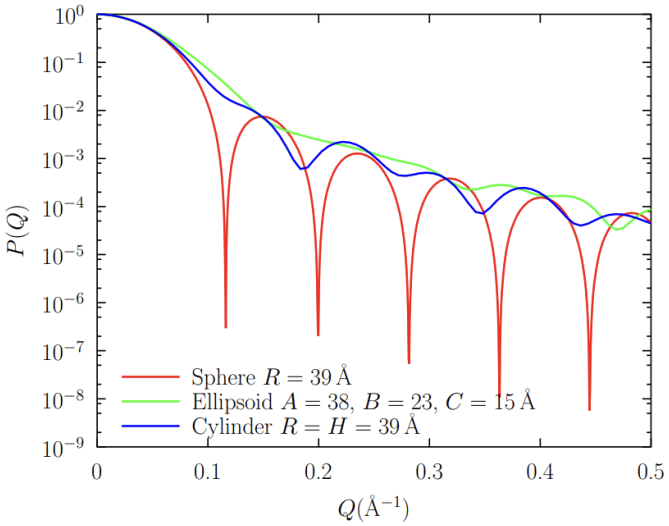


Figure 3.21 : Plot of P(Q) of model particles.

It can be observed that the three curves at very low Q value ($Q \rightarrow 0$) have a similar behavior. This means that the three shapes have the same size. Indeed, the low Q region contains information only about the size, represented, for SAXS, by the radius of gyration of the particle, R_g. This parameter, for homogeneous particles, is defined by the average of the squared distance of any point inside the particle from its geometrical center (*Eq. 14*).

$$R_g^2 = \frac{1}{V_p} \int_{V_p} r^2 d\vec{r} \qquad (14)$$

The importance of R_g is based on the validity of the law of Guinier, which is a simple approximation of the Taylor series expansion of $P(Q)$ according to (*Eq. 15*) :

$$P(Q) = 1 - \frac{Q^2 R_g^2}{3} + \cdots \approx e^{-\frac{Q^2 R_g^2}{3}} \qquad (15)$$

If law of Guinier lead to the approximation at low Q, the low of Porod defines the approximation of $P(Q)$ at large values of Q. This law shows a Q^{-4} power behavior and depends on the particle surface S_p, *Eq. 16.*

$$P(Q) \approx \frac{2\pi S_p}{Q^4 V_p^2} \qquad (16)$$

Figure 3.22 shows the relation between Guinier and Porod for a same particle and clearly shows the goodness of the two approximation at small and large Q, respectively.

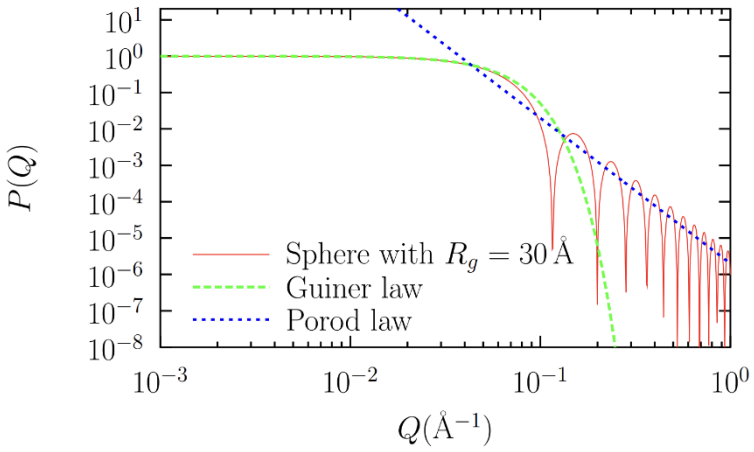


Figure 3.22 : Guinier (green line) and Porod (blue line) approximations of the form factor of a sphere.

Another useful representation of SAXS curves is the Kratky plot, which gives an overall information about the globularity of proteins in solution. This plot reports Q on x-axis and $Q^2\frac{d\Sigma}{d\Omega}(Q)$ on y-axis. It is sufficient to look at the plot to understand the general protein configuration. It helps to know if protein is folded and with a globular structure or if it is totally of partially unfolded. As can be seen in *Figure 3.23*, if the Kratky plot shows a main peak, the protein is in a compact state, instead a curve that reaches a plateau, eventually with small

negative or positive deviations, indicates unfolded or partially unfolded protein states.

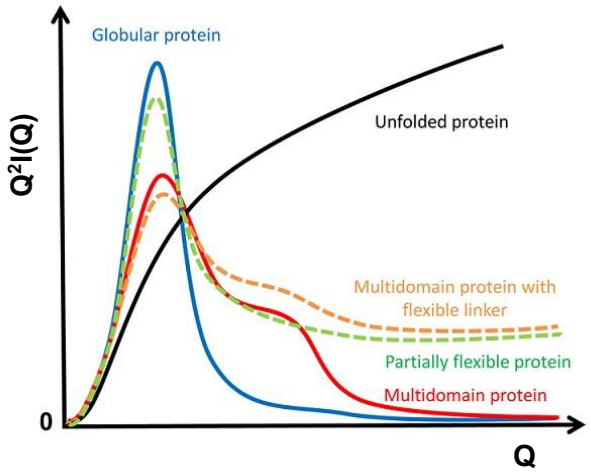


Figure 3.23 : Example of Kratky plot of proteins with different structures.

3.6.4 Experiment set up

We have investigated the protein ACE2 with SAXS technique, in the absence and in the presence of either modified sugar EC-202 or the modified sugar EC-312, at different temperatures from 20°C to 50°C. Measurements were made with highly automatized instrument at the BM29 beamline of the European synchrotron ESRF (Grenoble, France). This instrument operates using a 96-well microplate where samples are entered into strips of 8 x 200 µl

as well as into 12 individual tubes of 1.5 ml, typically used for the buffers (*Figure 3.24*) [32].

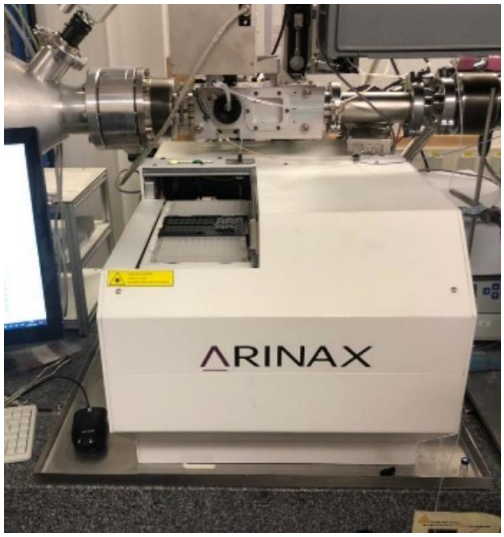


Figure 3.24 : Instrument in BM29. Samples are inserted in the comportment on the left.

With computer located in control cabin it is possible to employ the specific software to set any parameters like temperature or which buffer to use for any sample. The instrument automatically injects in the capillary the buffer, the protein sample and again the buffer (*Figure 3.25*). Ten measurements are taken for each solution passing in the capillary. One can observe the capillary throughout the experiment thanks to camera located above the capillary (*Figure 3.26A; Figure 3.26B*).

Figure 3.25 : Capillary with its support.

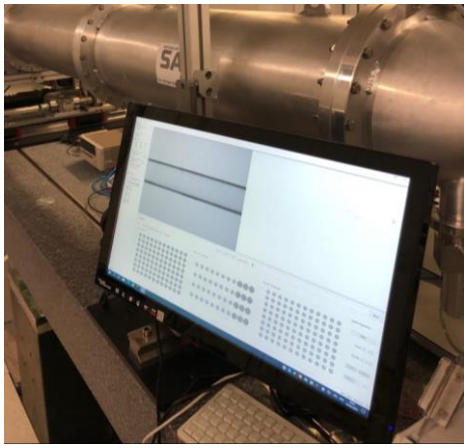


Figure 3.26A : Monitor to follow what happens in the capillary.

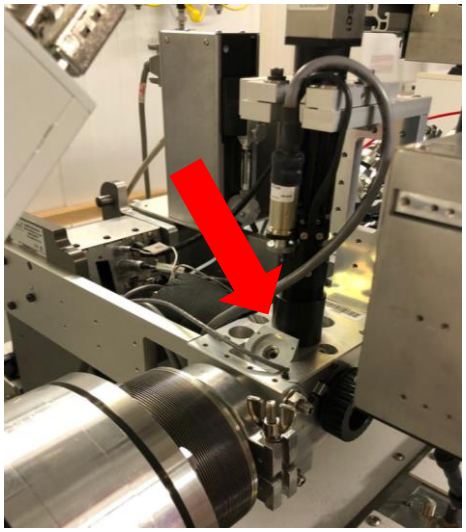


Figure 3.26B : Red arrow shows position of the camera.

3.7 Data analysis

3.7.1 DLS data

Malvern's software automatically analyses the correlation function using a CONTIN-based algorithm to calculate the size distribution of the particles in solution. However, the user has no control on this analysis and when the calculation is not able to reproduce the experimental data it is impossible to get the correct results. For this reason, we have reanalyzed all experimental correlograms with an in-house software application called DynaLiSc. It allows to upload data, cut noisy sections of the correlograms and calculate the size distribution histogram. The calculation method is based on CONTIN and uses a single free parameter, called factor α [33], defining the minimum width of a peak in the distribution. The smaller the α value, the narrower is the peak. This parameter can be changed to resolve multiple peaks that are partially overlapping. At the end of the process the intensity-weighted size distribution histogram can be downloaded in a text format to be further analyzed with any other software application.

In our analysis, for a better comparison between DLS and SAXS results and to have a more representative data that provide information about the amount of each type of particles in solution, it is preferable to work with volume-weighted size distributions instead of intensity-weighted ones. This

conversion was performed assuming the intensity-volume relation described in the framework of Rayleigh scattering.

3.7.2 SAXS data

SAXS data are made up 10 curves for each sample and relative buffer. The first think to do is to control if these curves are overlapped and to eliminate the ones which are not. This because we need to take the average of each curve, if just one of the 10 curves is different from the other, the average results incorrect. Plotting in the same graph the sample – empty cell and the buffer – empty cell one must see that sample curve always passes over the one of buffer. If this does not happens there is a problem in the data. Some of our data had this type of issue, but we fix it using a correction factor. With the right data we derived the subtraction between sample and buffer to obtain the proteins signal only and this was analyzed, also by applying Guinier law and Kratky plot to try to understand the protein configuration.

Chapter IV

RESULTS

4.1 DLS results

DLS experiments on two different proteins associated with four types of sugars at growing concentrations and at different temperatures were performed. For each sample, 27 measurements were taken, three for each temperature. Using software tools, data were analysed and using some scripts developed by us under Gnuplot software, they were processed to derive information about protein aggregation. At first, the DLS correlograms and the resulting intensity weighted size distribution $I(R)$ were plotted. Then, as described in Chapter III, from the average of the three intensity weighted size distribution measurements, the volume fraction distribution $\varphi(R)$ was derived. This allows to calculated the relative quantity of particles according to their size. According to *Eq. 5*, the volume distributions were fitted for each temperature by a sum of five Gaussian curves.

We have prepared many protein samples which have been measured with DLS and therefore many plots have been obtained. For this reason we have decided to report the analysis with Gaussian only for few representative cases. We show, for each sample, the volume fraction distribution $\varphi(R)$ from which

the main results, that are the fraction of monomers forming the k-oligomers (see *Eq. 6*) were obtained. Furthermore, in this chapter we show results for one sugar concentration for each sugar. Results with other sugars are reported in the *Appendix*.

4.1.1 Myoglobin's results

First of all, results of myoglobin without sugars, which is the control sample, are shown. In this case we include all the plots, with the aim of showing in detail the performed analysis.

The first plot is the DLS correlogram. It represents all the DLS measurements with three repetition for each temperature, together with the best fit for performed with the software DynaLiSc. We treated the intensity weighted size distributions $I(R)$ with Gnuplot scripts written by us, to convert theme in volume weighted size distributions $V(R)$, to normalize the resulted curves by calculating the area under the curve with trapeziodal method (see Chapter III for details), to perform the average of the three measurements for each temperature and, finally, to plot the derived volume fraction distributions $\varphi(R)$. Subsequently, we performed a fit of the volume fraction curve with a sum of five Gaussians, according to *Eq. 5*. We carefully selected the validity range of the peak centers R_i values, considering the hypothetical radius of oligomers. At the end, the resulted parameters of Gaussian curves were

interpreted in relation to the particles size calculated by Piccinini et al. (2022). In particular for myoglobin there are three possible conformation, monomer with an average radius of 17.0 ± 0.2 Å, dimer with radius 26.7 ± 0.3 Å and unfolded monomer, with radius of 43 ± 2 Å. Considering these values and the positions of the five Gaussian curves we interpreted the first three Gaussian as representative of the monomer, the fourth representative of the dimer and the fifth one of the unfolded conformation. On the basis of *Eq. 6*, fraction of the monomers distributed in each protein state has been calculated and plotted in the last graph.

Hereafter, regarding myoglobin without sugar samples, plots of correlograms (*Figure 4.1A*), intensity weighted size distributions derived from correlograms (*Figure 4.1B*), average volume fraction distribution over three repeated experiments (*Figure 4.1C*), analysis with Gaussians curves (*Figure 4.1D*) and oligomer distribution (*Figure 4.1E*) are shown.

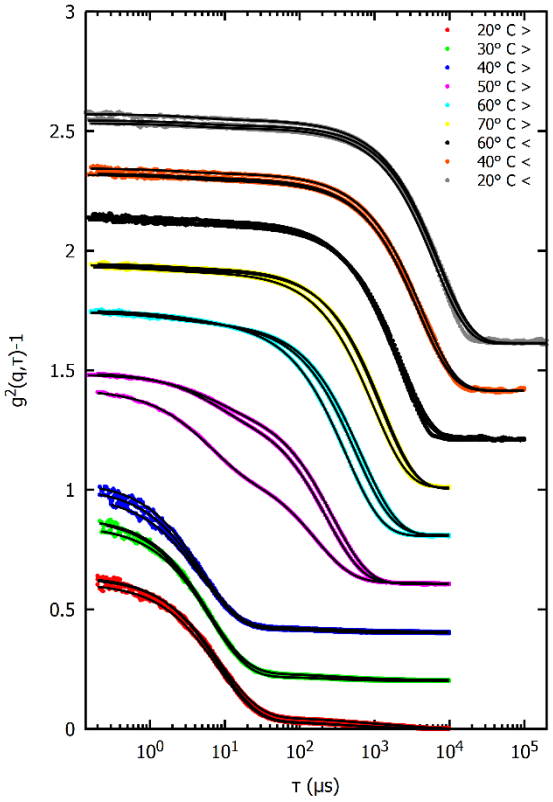


Figure 4.1A: DLS Correlograms of myoglobin without sugars. Three measurements for each temperature with their relative fit are reported.

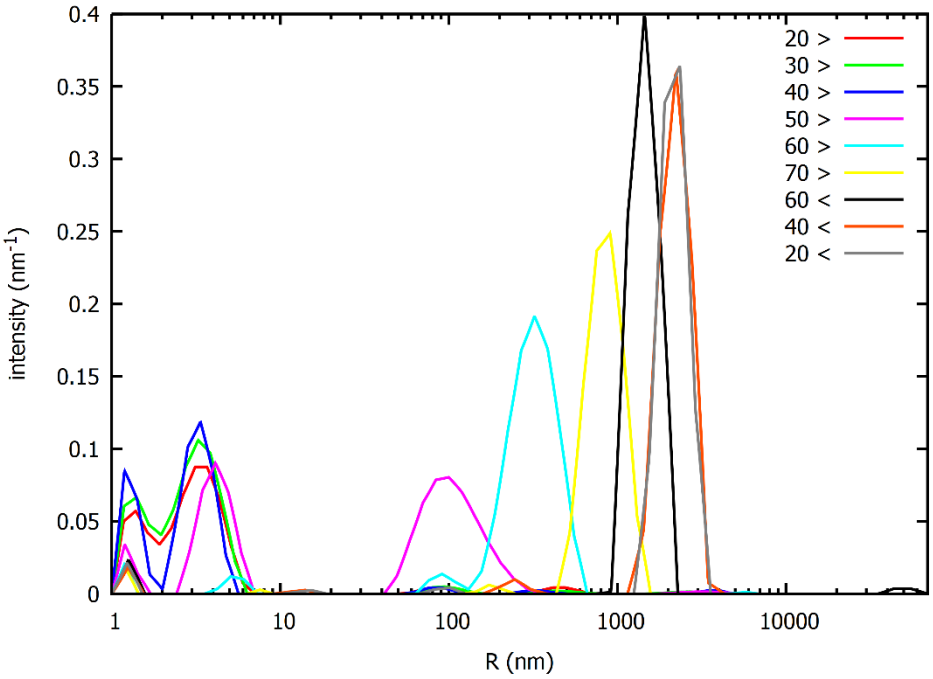


Figure 4.1B : Plot of intensity distribution. In this case just a measure for each temperature is represented.

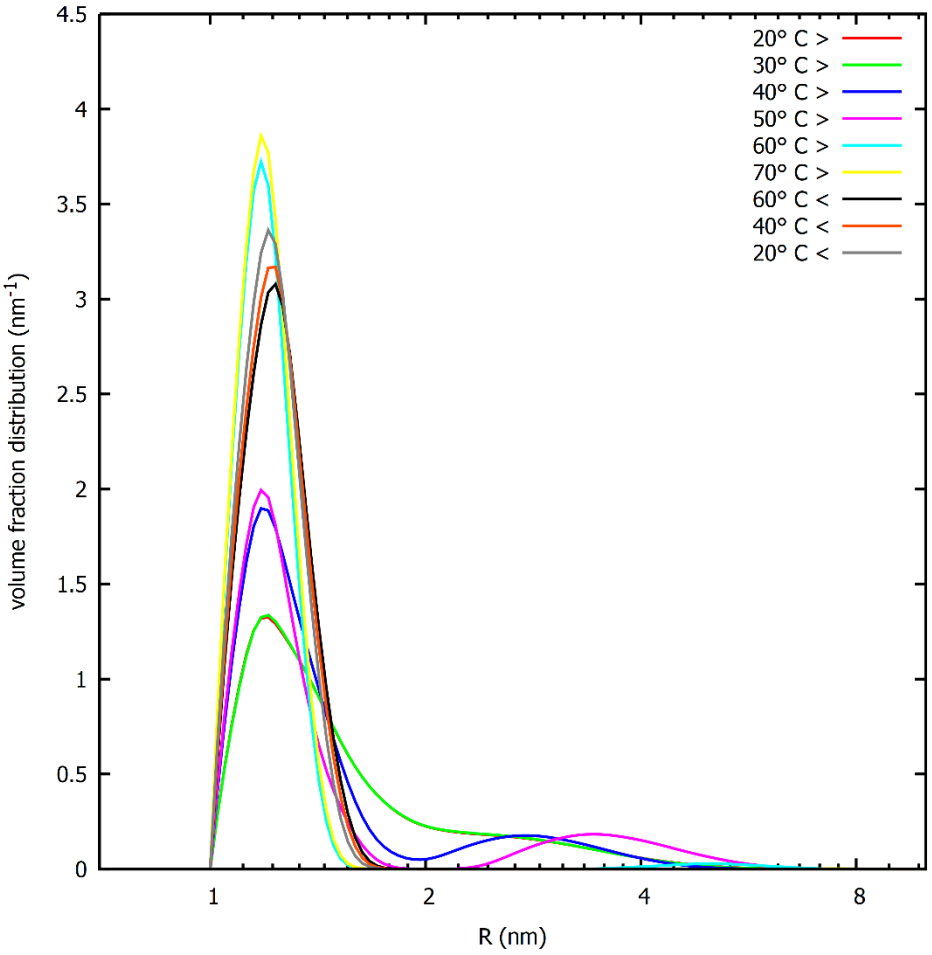


Figure 4.1C : Volume fraction distribution.

In the intensity weighted size distribution plot, it is possible to detect the presence of very large particles, some of them with size greater than 1000 nm. Looking at the volume fraction distribution plot this very large particles there are not visible. These particles give high signal, which cover the other signals, in the intensity weighted size distribution plot just because they are much bigger than the other particles in solution but they are very low in number or volume weighted size distribution ($N(R)$ or $V(R)$, respectively). For this

reason, during the conversion in the volume distribution they have not be considered.

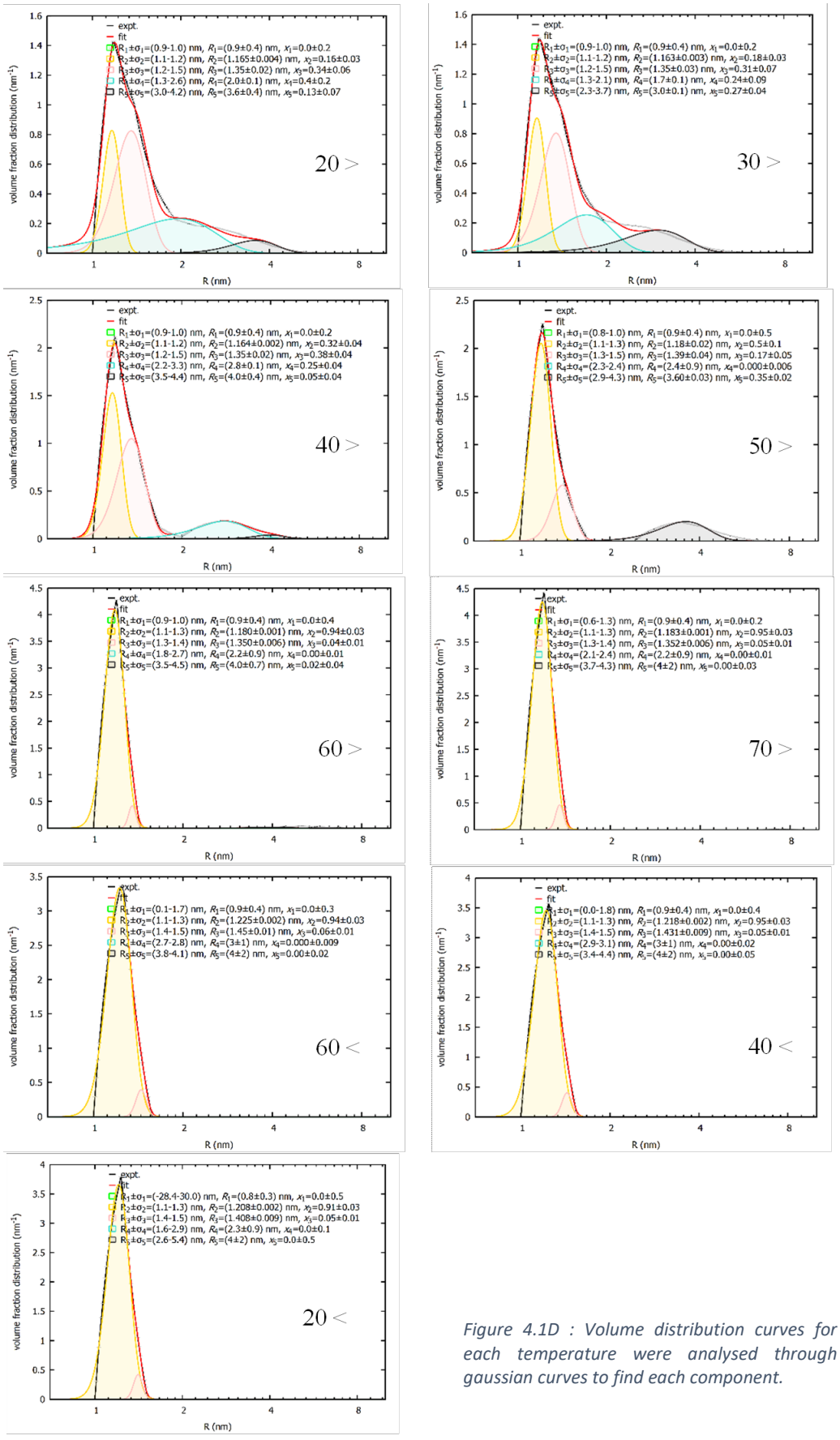


Figure 4.1D : Volume distribution curves for each temperature were analysed through gaussian curves to find each component.

We decided to report all these type of plots only for this sample, but this analysis was carried out for all samples. In the plots above there are the value of σ (sigma) which is the amplitude of the Gaussian curve and the calculated radius of particles, whit its error, for each contribution. These values were used to make the plot below, which represents the molar fraction of monomers which form the different conformations.

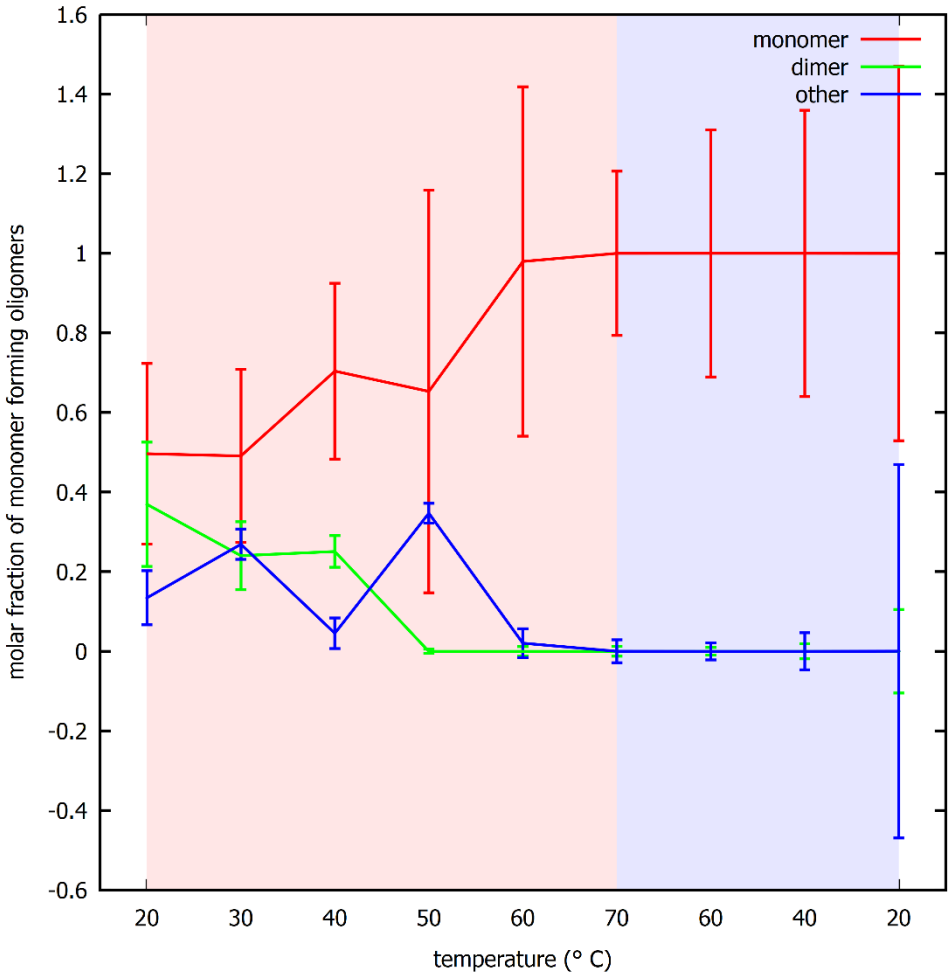


Figure 4.1E : Oligomer distribution of myoglobin monomers without sugars as a function of temperature.

In this situation all different conformations, in different fraction, are present at the beginning, but when temperature increases, and also decreases, all the protein are in the monomeric state.

For the other samples, with sugar at concentration of 0.25 M, we report only the volume fraction distribution plot and the one of molar fraction of monomer.

Myoglobin 2 g/L + EC-101 0.25 M

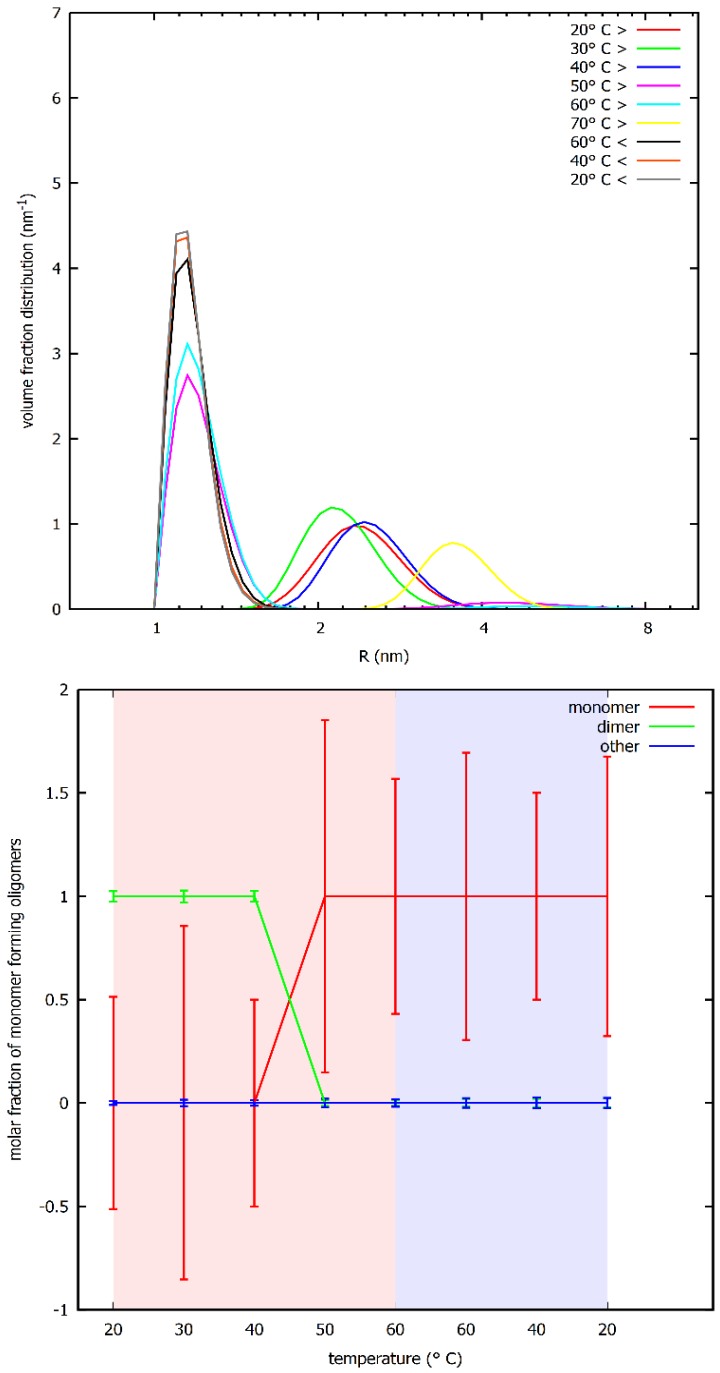


Figure 4.2 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 2 g/L + EC-101 0.25M.

In the second plot there is not represent the situation at 70°C because it was difficult to find the Gaussian curve for the respective volume fraction curve.

Myoglobin 2 g/L + EC-202 0.25 M

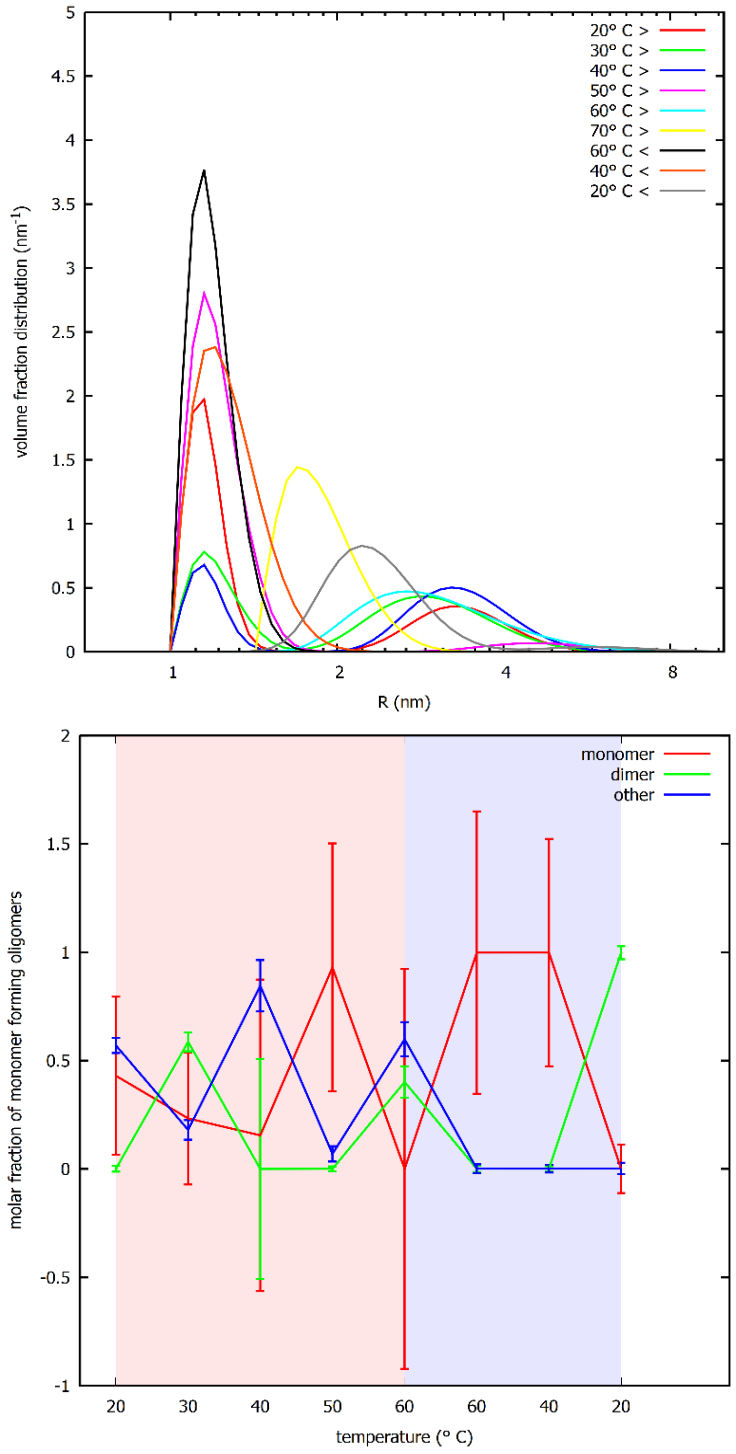


Figure 4.3 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 2 g/L + EC-202 0.25M.

Also in this case the measure at 70°C was discarded.

Myoglobin 2 g/L + EC-212 0.25 M

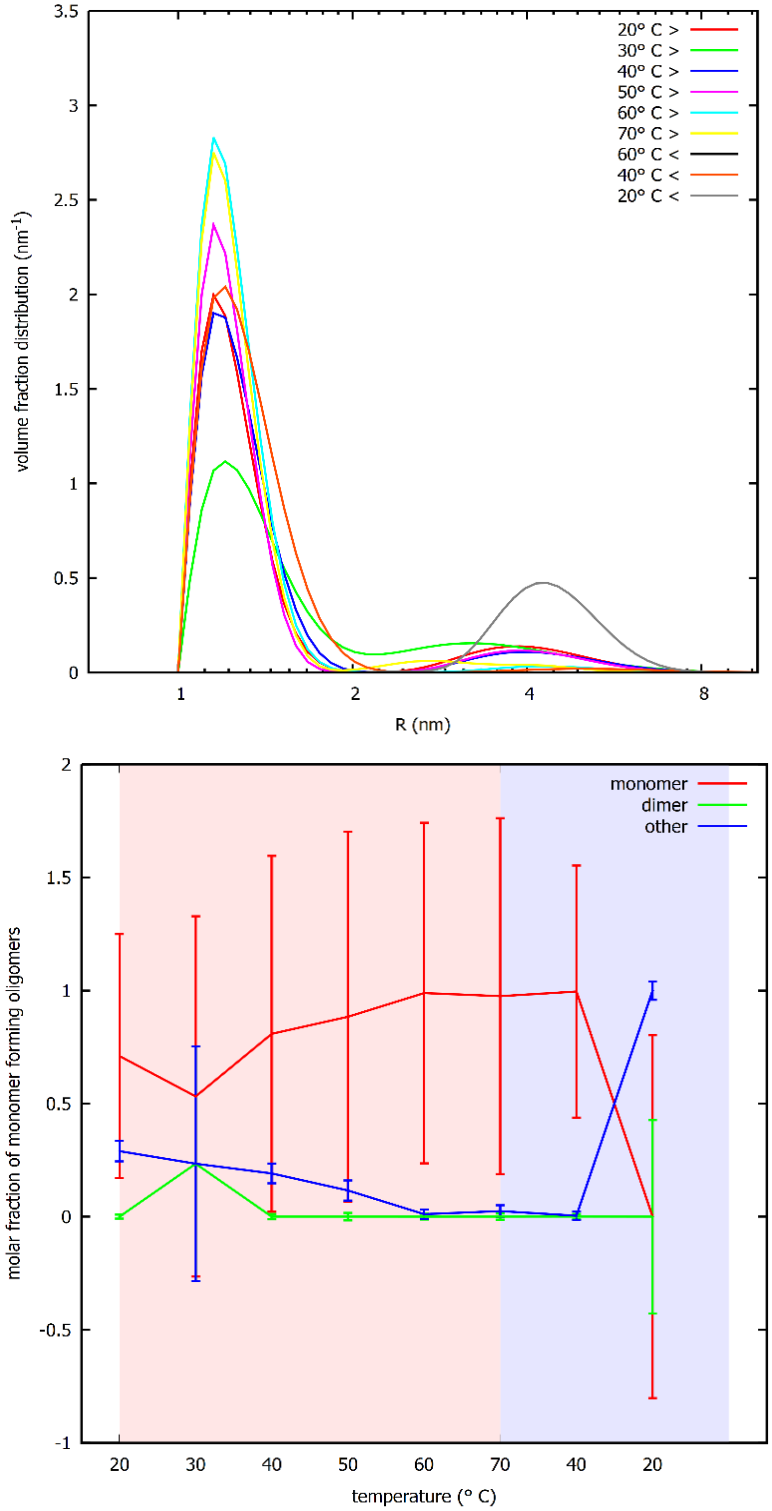


Figure 4.4 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 2 g/L + EC-212 0.25M.

Myoglobin 2 g/L + EC-312 0.25 M

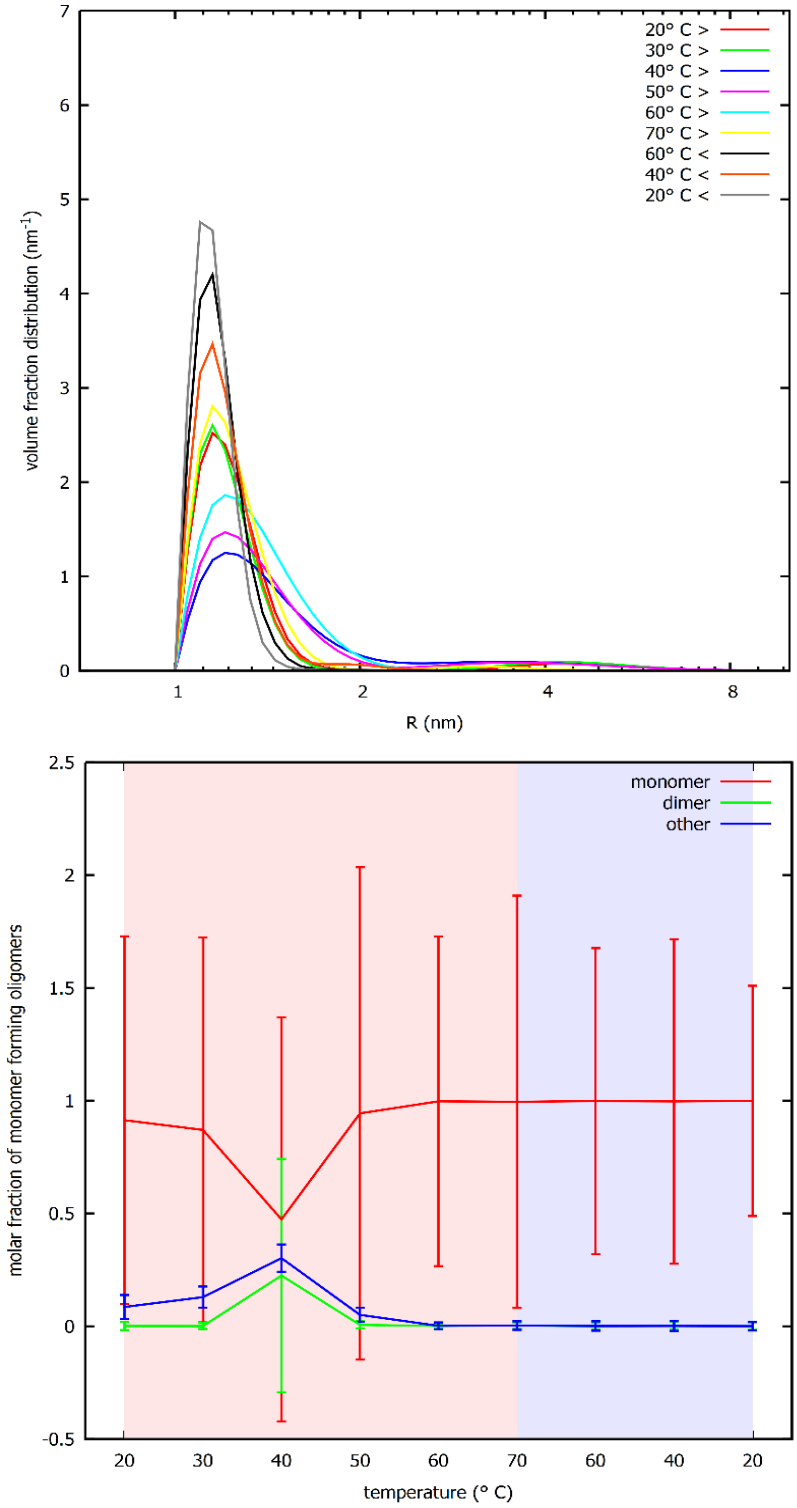


Figure 4.5 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 2 g/L + EC-312 0.25M.

Above the results for myoglobin at 2 g/L in association with each sugar at their highest concentration is reported. In the plots of molar fraction of monomer it is possible to see the general trend of the abundance of the different conformations. The lack of a measure in some cases does not affect the general trend. We can analyse the behaviour of protein in relation to different sugars. Sugars EC-101, EC-212 and EC-312 present similar trend and the most stable situation seem to be the EC-312, which presents almost exclusively monomer (*Figure 4.5*). With EC-202 there is a similar situation but dimer and unfolded structure are more represented (*Figure 4.4*). In solution with EC-101, myoglobin is mostly in dimeric conformation at the beginning and subsequently only monomers are present, with absence of unfolded form (*Figure 4.2*). However myoglobin with EC-212 has a unusual behaviour, with continuous changes of conformation and it is not possible to notice a specific trend (*Figure 4.3*).

Then we report myoglobin at 4 g/L always with each sugar at the higher concentration.

Myoglobin 4 g/L + EC-101 0.25 M

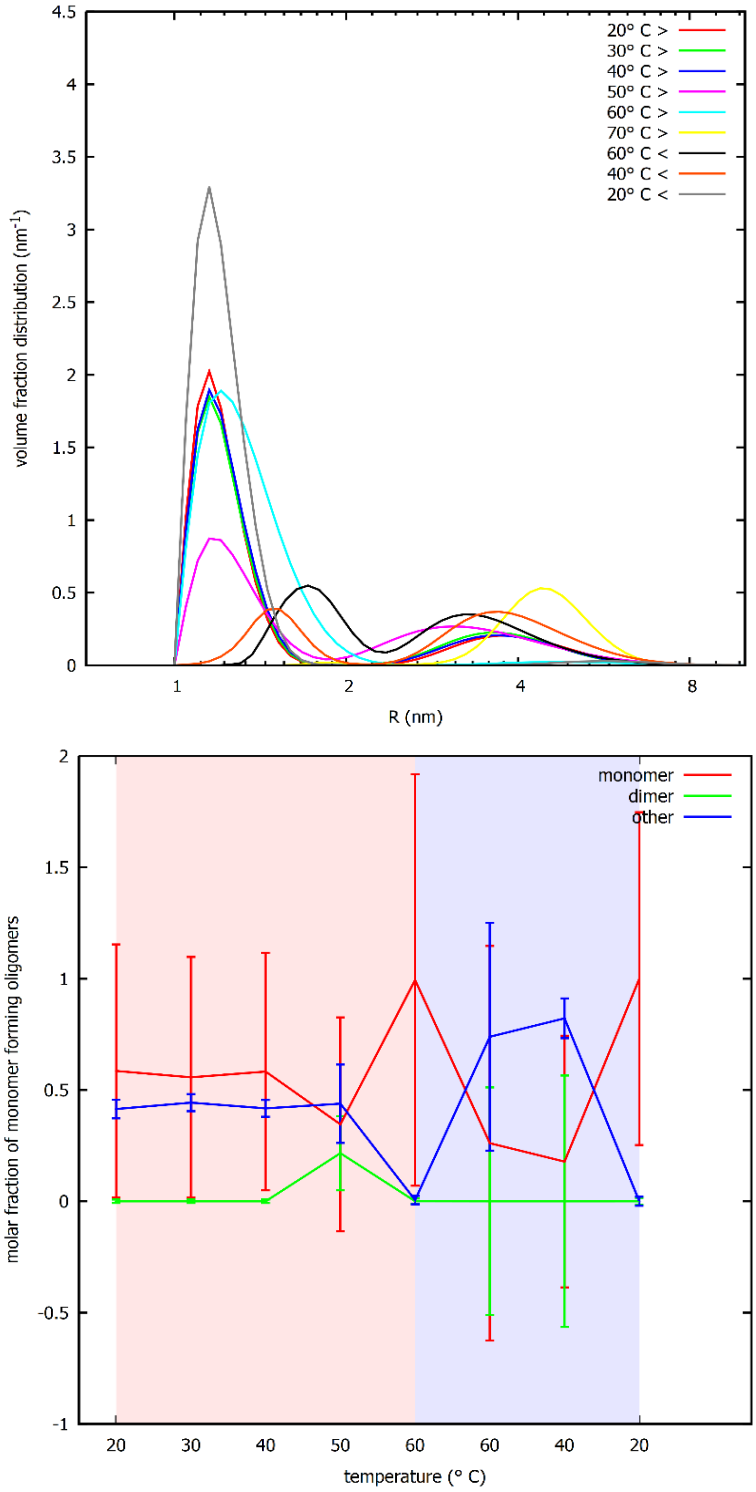


Figure 4.6 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 4 g/L + EC-101 0.25M.

Myoglobin 4 g/L + EC-202 0.25 M

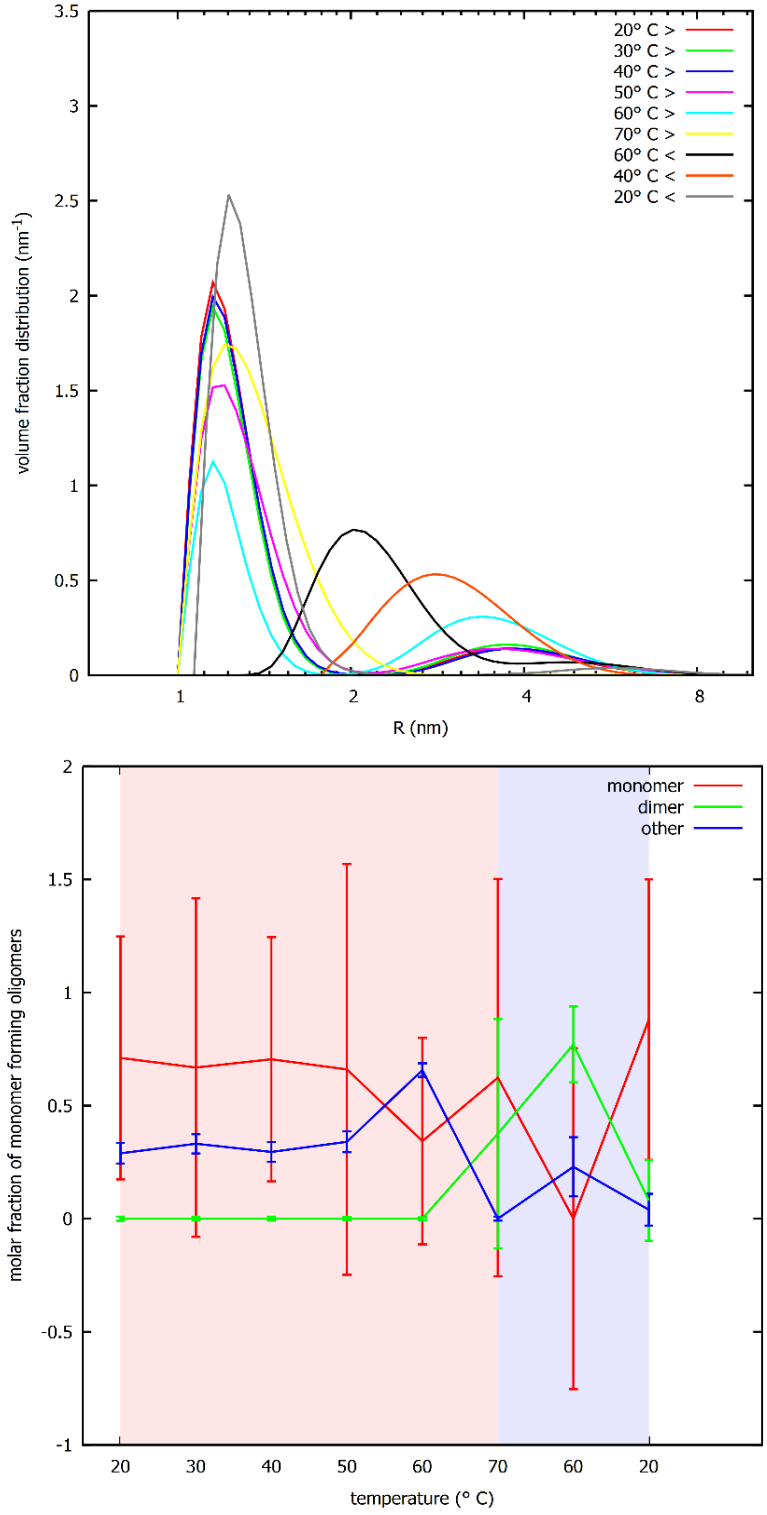


Figure 4.7 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 4 g/L + EC-202 0.25M.

Myoglobin 4 g/L + EC-212 0.25 M

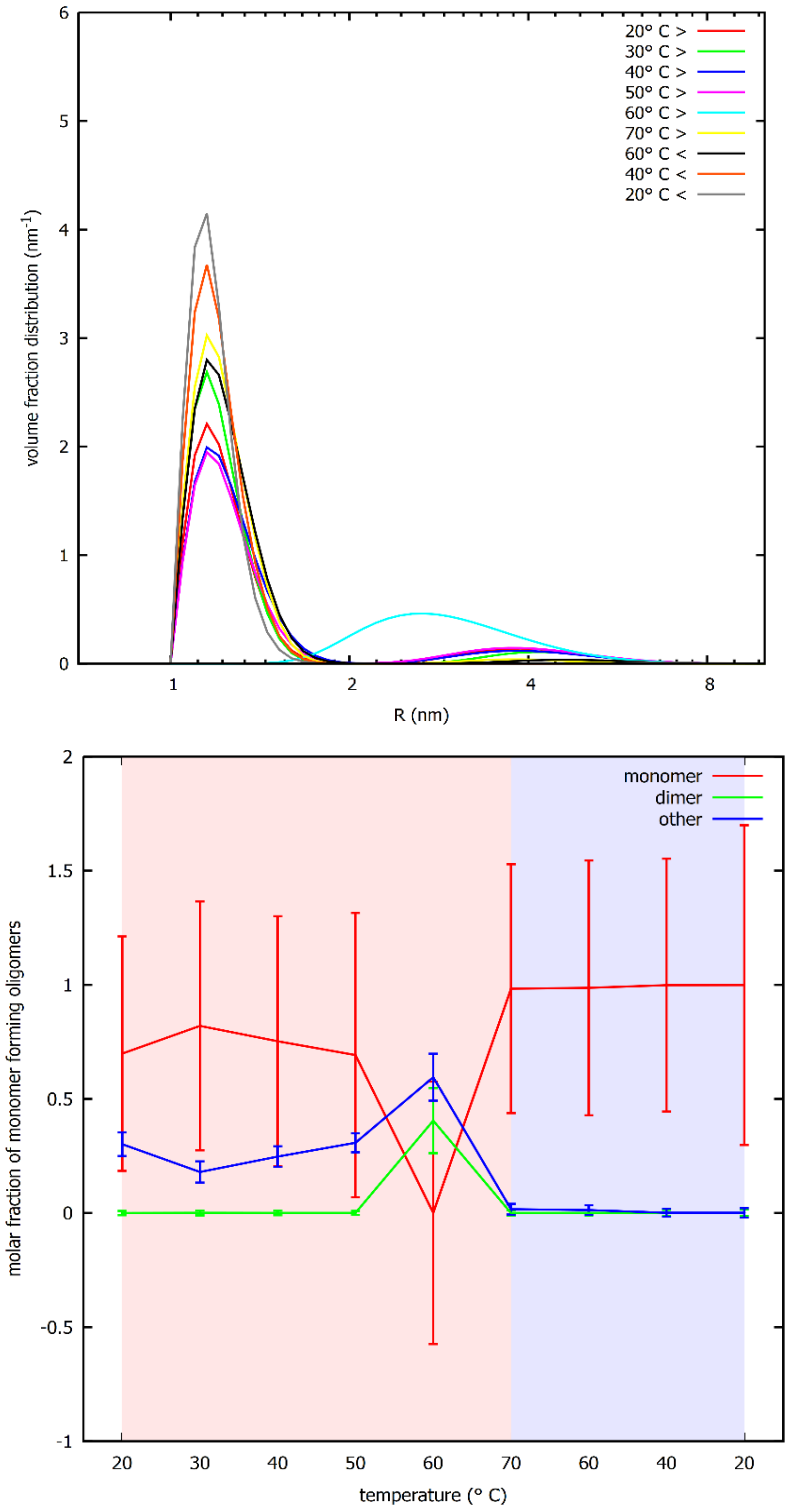


Figure 4.8 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 4 g/L + EC-212 0.25M.

Myoglobin 4 g/L + EC-312 0.25 M

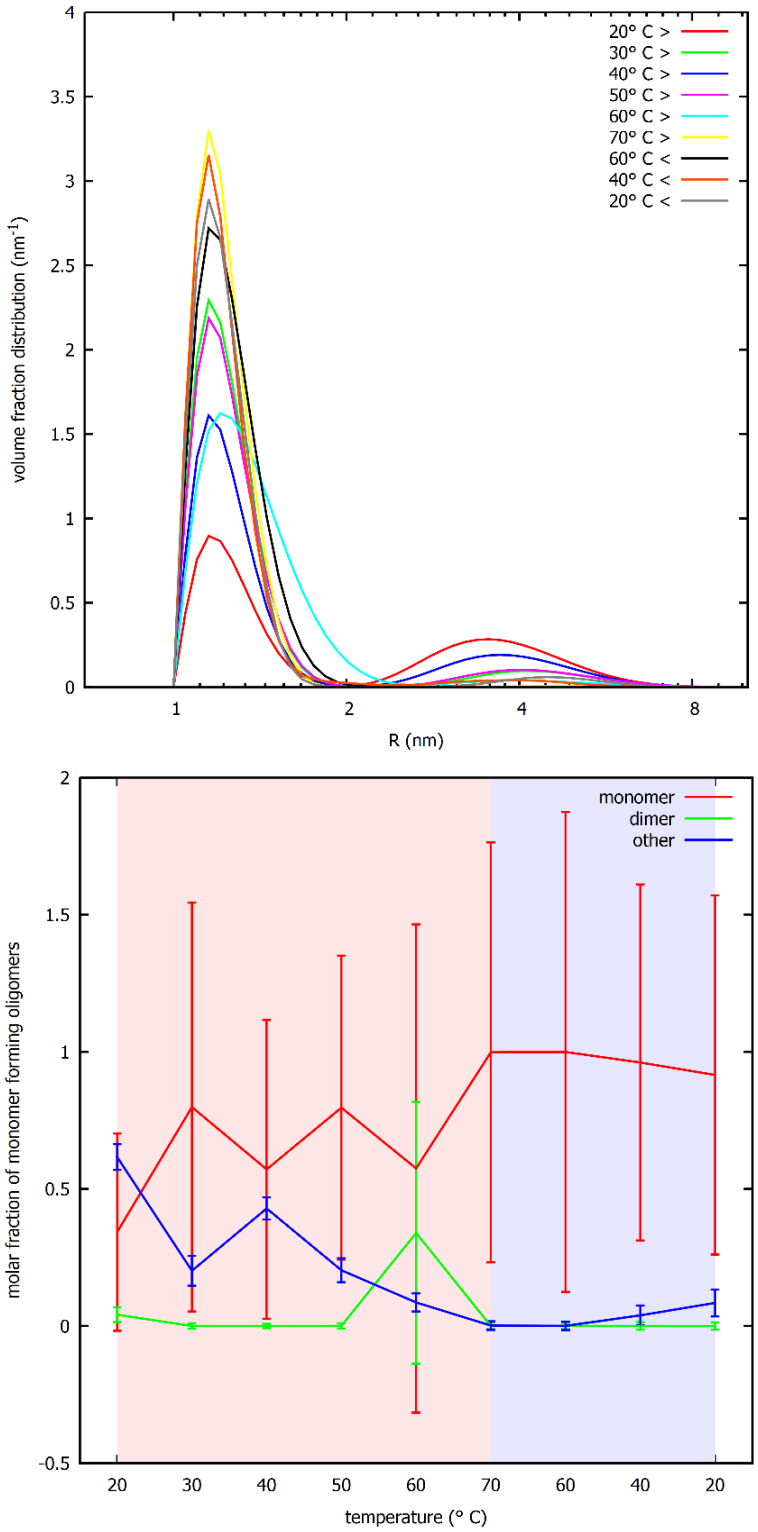


Figure 4.9 : Volume fraction distribution and molar fraction of monomer plots for myoglobin 4 g/L + EC-312 0.25M.

The trend of each plot with myoglobin at 4 g/L is not so defined compared to the same situation with protein at 2 g/L. As in the previous analysis, the best sugars to stabilize protein result EC-312 and then EC-212, but here there are increasing amounts of dimers and unfolded conformations (*Figure 4.9; Figure 4.8*). With EC-101 and EC-202 the behaviour is similar, at the beginning, until 50°C, there are more monomers and less unfolded states and there are not dimers, then the situation turns in a not well-defined trend. For samples with EC-101, the result is completely different compared to the one with myoglobin at 2 g/L (*Figure 4.6*). Also the sample with EC-202 at the beginning shows a more ordinated behaviour than the same situation with protein at 2 g/L and then they become very similar.

4.1.2 Insulin's results

Also we report the results for insulin in the same way. Insulin was analysed likewise myoglobin, first a fit using software DynaLiSc was performed on correlograms and the intensity distribution was derived. Then on the intensity distribution several Gnuplot scripts to make average, to convert in volume fraction and to normalize it were applied. Also the process with Gaussian curves was performed to find the contribution of each particle conformation based on the size and then these results were plotted to understand the molar fraction of monomer distributed in each conformation.

Insulin forms five different conformations, calculated by Piccinini et al. (2022), which are monomers, with radius of 9.6 ± 0.1 Å, dimers, with radius 13.2 ± 0.5 Å, tetramers, with radius 23.3 ± 02 Å, hexamer, with radius 27.0 ± 0.3 Å, and unfolded conformation with larger size.

We only report the volume distribution (*Figure 4.10*) and the molar fraction of monomer (*Figure 4.11*) plots. Also for this protein only the result for each sugar at 0.25 M are shown.

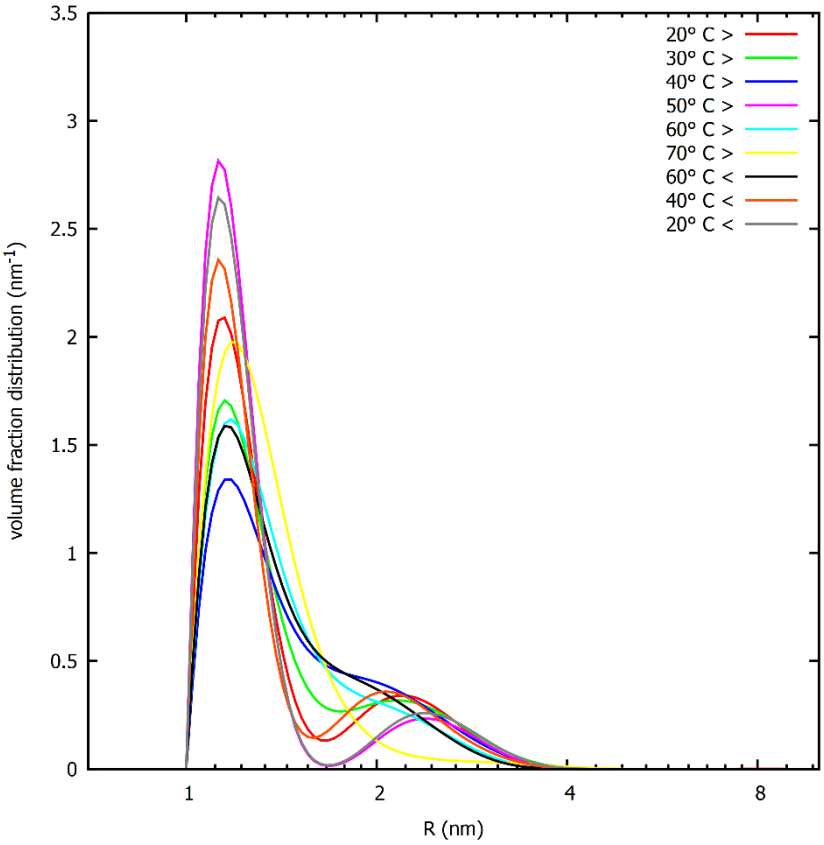


Figure 4.10 : Volume fraction distribution plot derived by conversion of the intensity distribution.

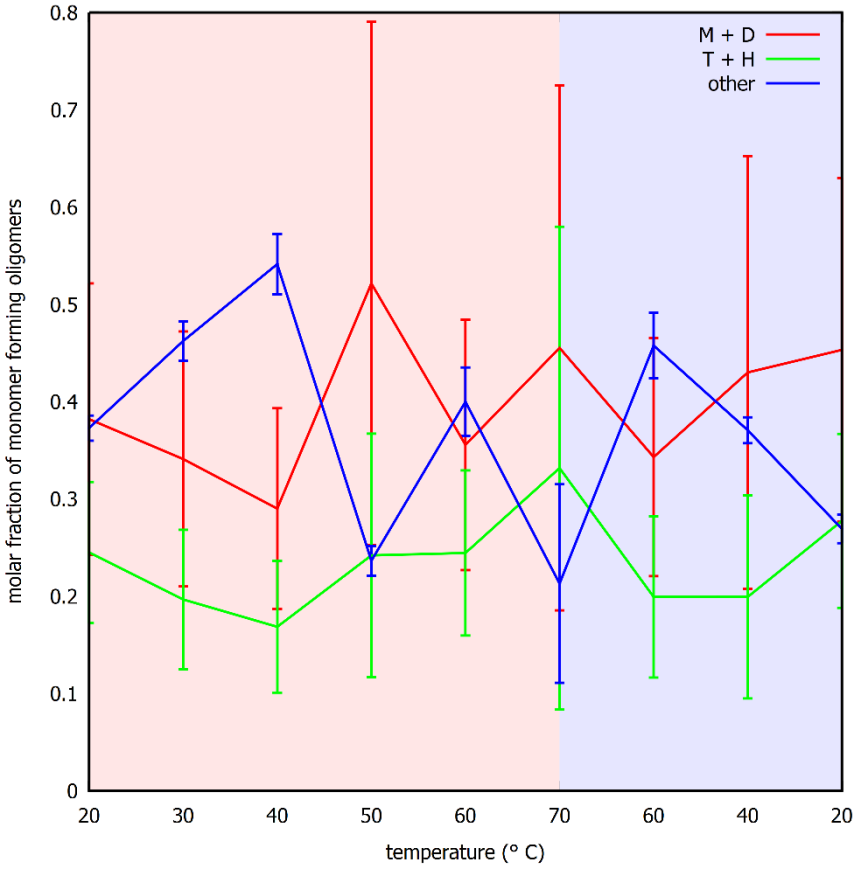


Figure 4.11 : molar fraction of monomer which represents the percentual amount of each conformation.

The red line represents the sum of monomer and dimer (M + D), the green line is tetramer and hexamer (T + H) and the blue line is particles of larger size. Insulin sample, without sugars, not presents a linear trend, which indicates that it is unstable by itself.

Below we show the situation in presence of sugars at 0.25 M.

Insulin 2 g/L + EC-101 0.25 M

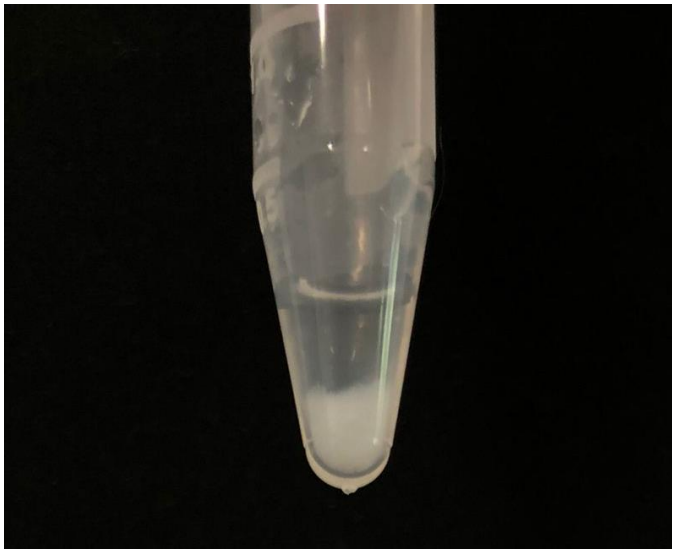


Figure 4.12 : Tube containing the sample with EC-101 at 0.25 M.

Insulin 2 g/L + EC-202 0.25 M

Figure 4.13 : Tube containing the sample with EC-202 at 0.25 M.

Insulin 2 g/L + EC-212 0.25 M

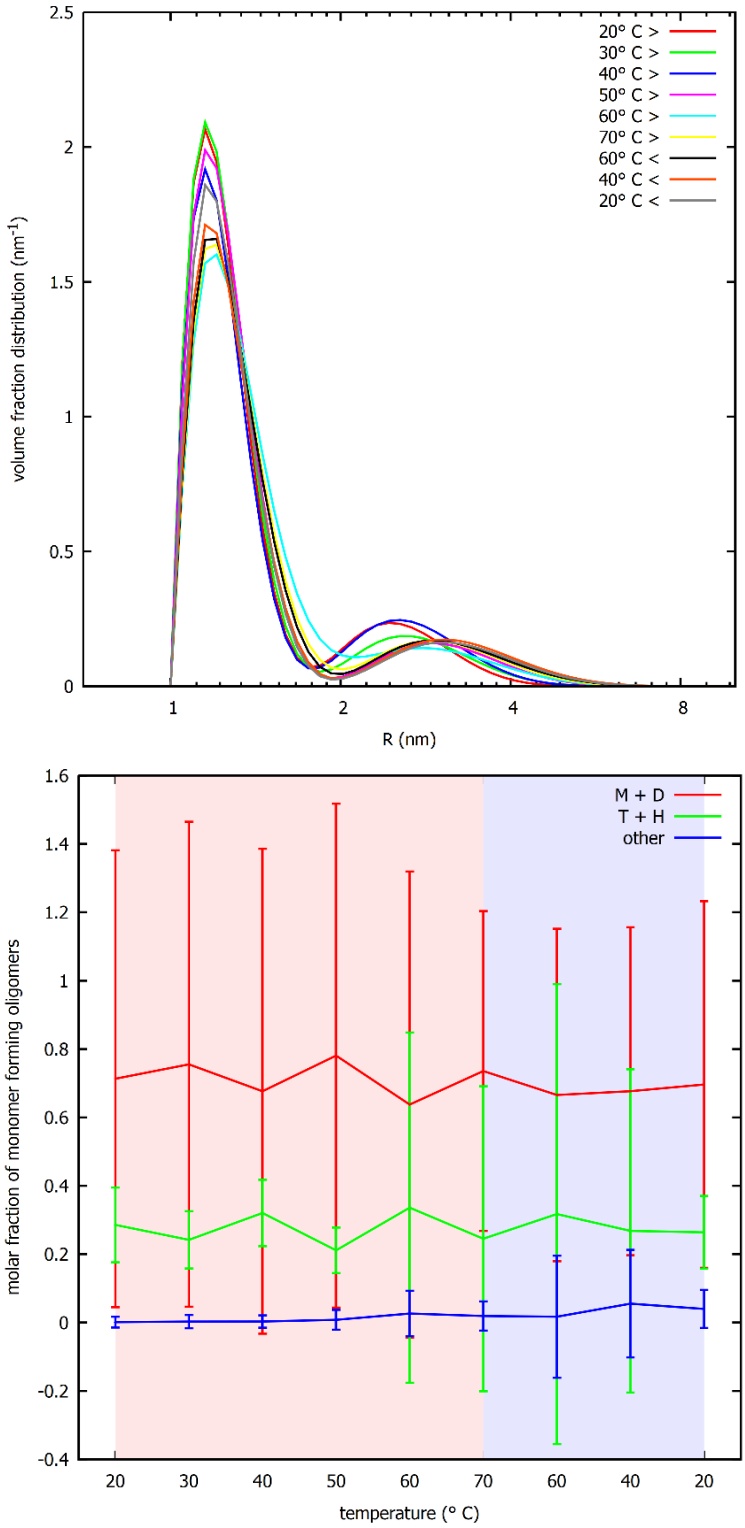


Figure 4.14 : Volume fraction distribution and molar fraction of monomer plots for insulin 2 g/L + EC-212 0.25M.

Insulin 2 g/L + EC-312 0.25 M

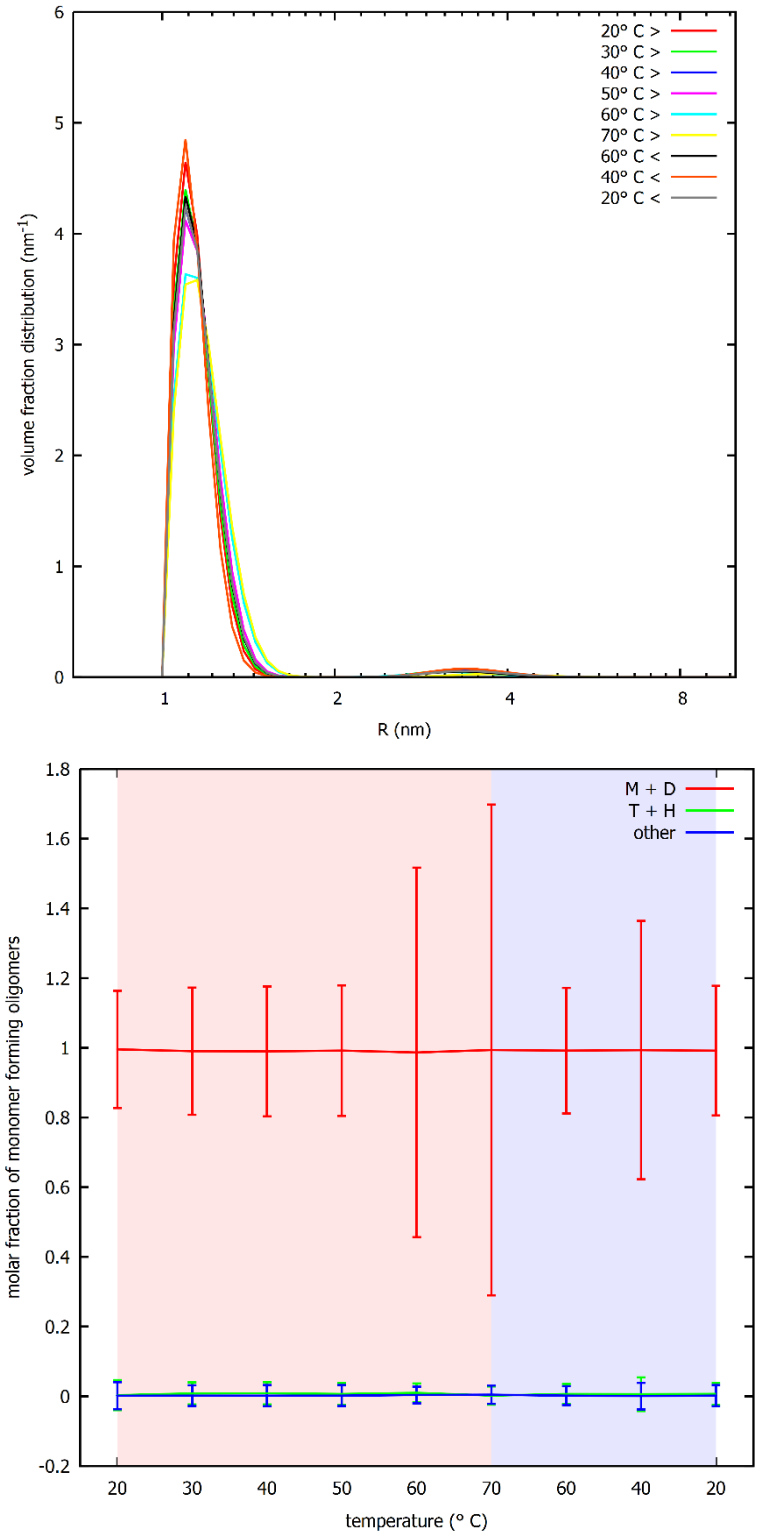


Figure 4.15 : Volume fraction distribution and molar fraction of monomer plots for insulin 2 g/L + EC-312 0.25M.

For insulin the situation is more defined than myoglobin. During samples preparation we noticed that when we added sugars EC-101 (*Figure 4.12*) and EC-202 (*Figure 4.13*) to the insulin the samples showed precipitates. This is a clear indication of aggregation and these two samples were not analysed because all protein was precipitated. For EC-212 (*Figure 4.14*) and EC-312 (*Figure 4.15*), the situation is completely different. In both cases, there is a large amount of monomer and dimer with differences for the other conformation. In fact with EC-212 insulin is also in tetramer and hexamer conformations but with EC-312 tetramer, hexamer and larger structures are totally absent.

4.2 Results of SAXS analysis

SAXS experiments were performed at different sugars concentration and by varying temperature. Data analysis was performed with the Gnuplot software, by using several scripts to elaborate results and to represent the curves on plot.

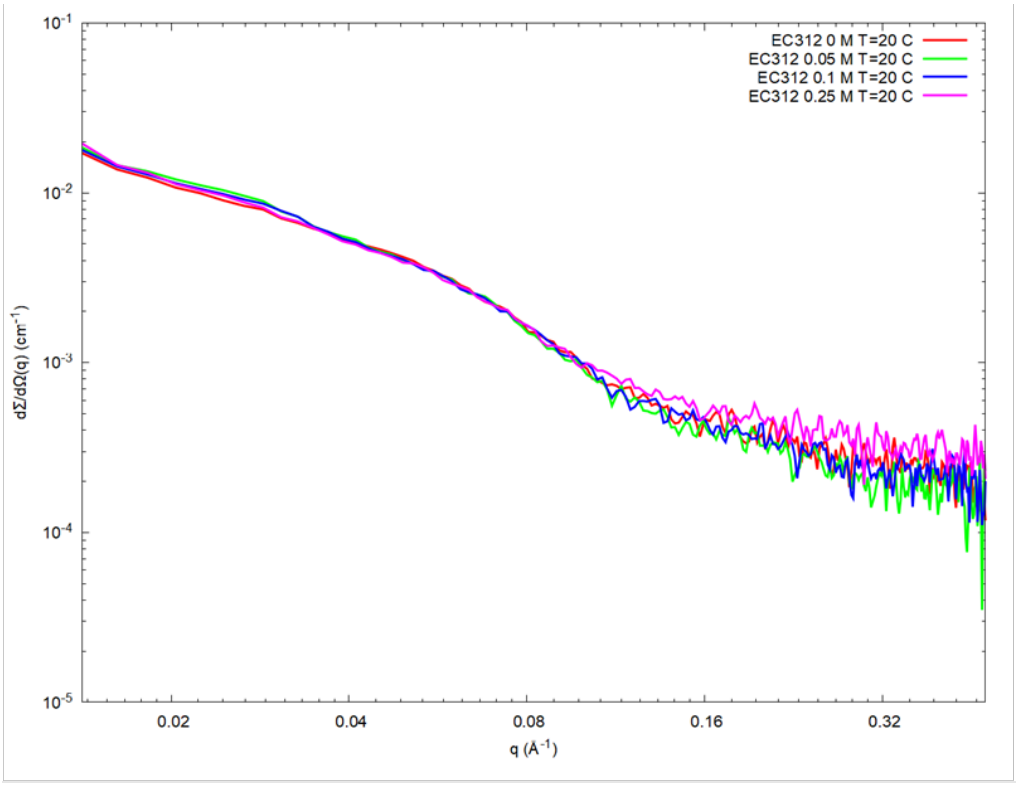
During SAXS experiments, ten SAXS measurements for each sample were recorded, but not all the repeated measurements were considered, because, in some cases, air bubble or dust could have passed through the

capillary. For this reason, it is important to check the similarity of the repeated curves and eventually to discard the ones that are clearly outlier. Then a curve average is made among all the remaining curves. SAXS curves of protein samples are obtained by subtracting the buffer signal.

4.2.1 SAXS curves

SAXS curves are reported in the following plots, which report the intensity I(Q), expressed in cm^{-1}, as a function of the modulus Q of the scattering vector, expressed in $Å^{-1}$.

Each plot reports SAXS curves at a fixed temperature (20°C, 30°C, 40°C, 50°C) and by varying the concentrations of one of the two sugars (0 M, 0.05 M, 0.1 M, 0.25 M) (*Figure 4.16A, B, C, D*).



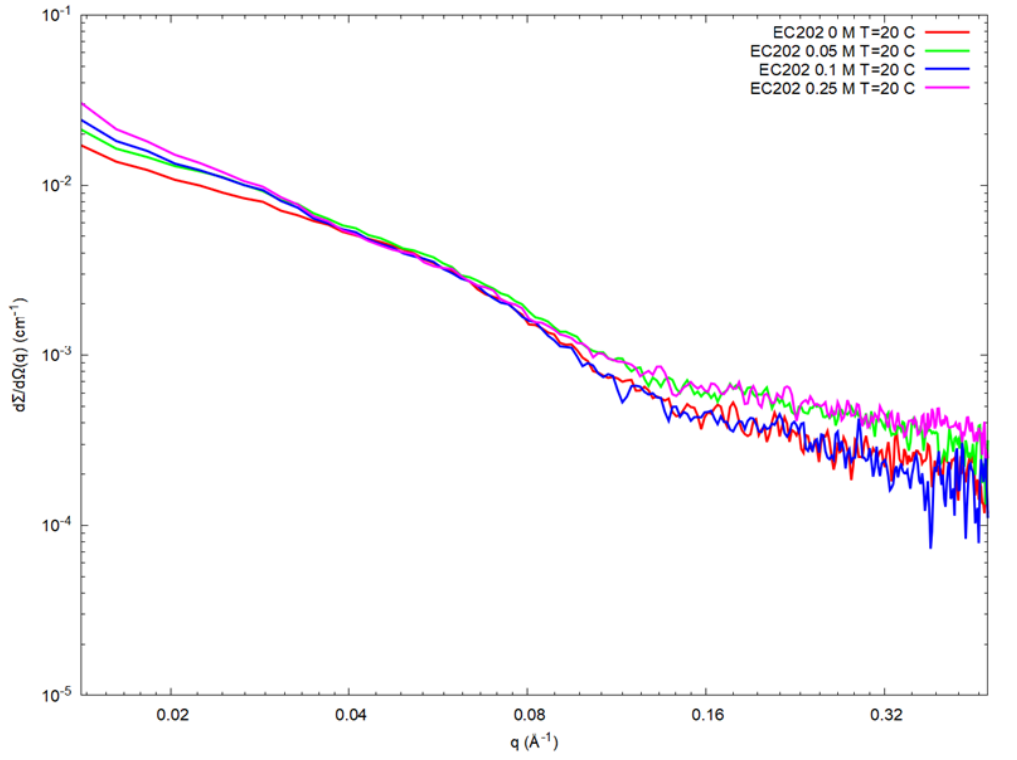


Figure 4.16A : SAXS plots which report ACE2 with sugars at 20°C.

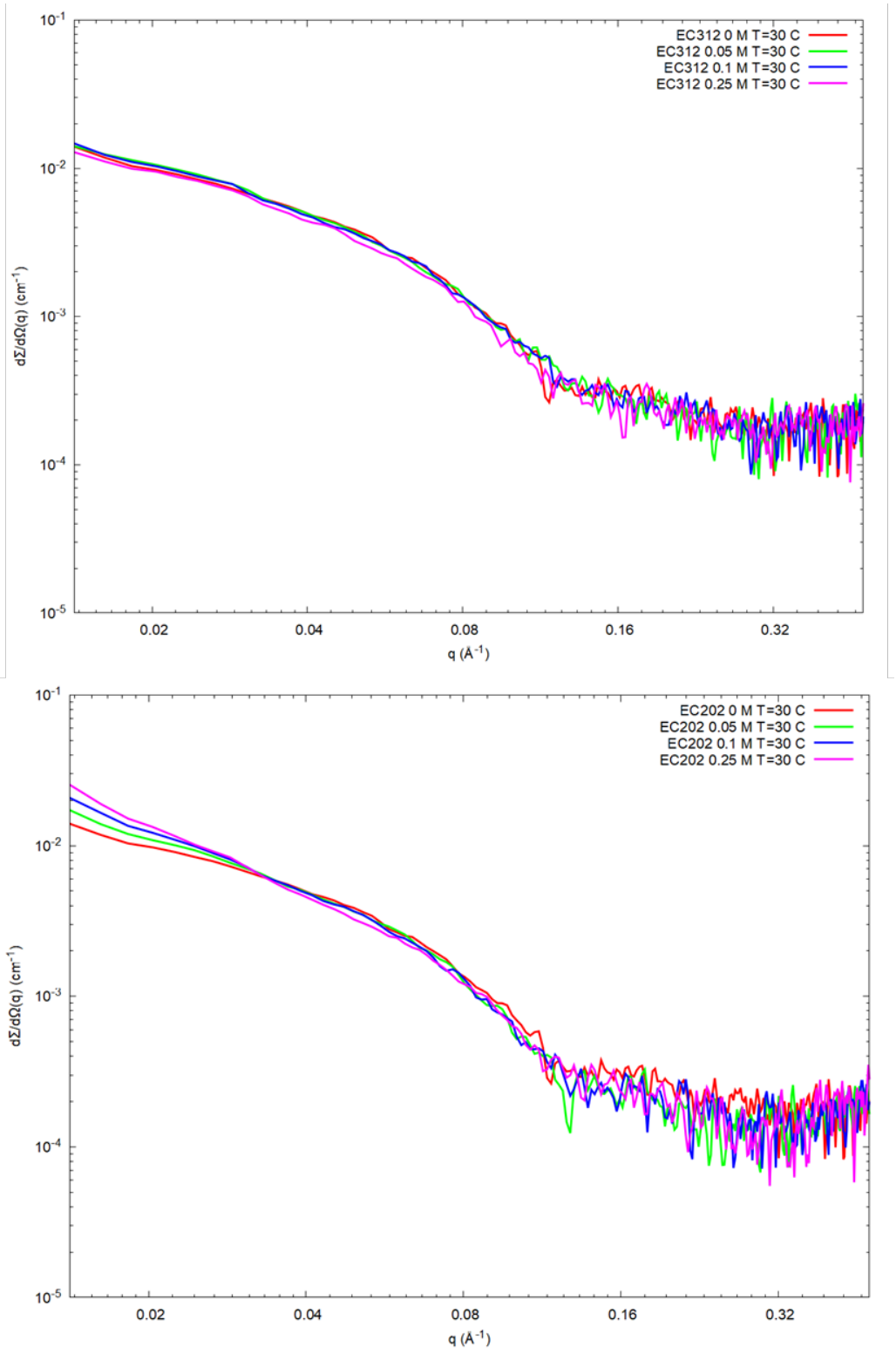
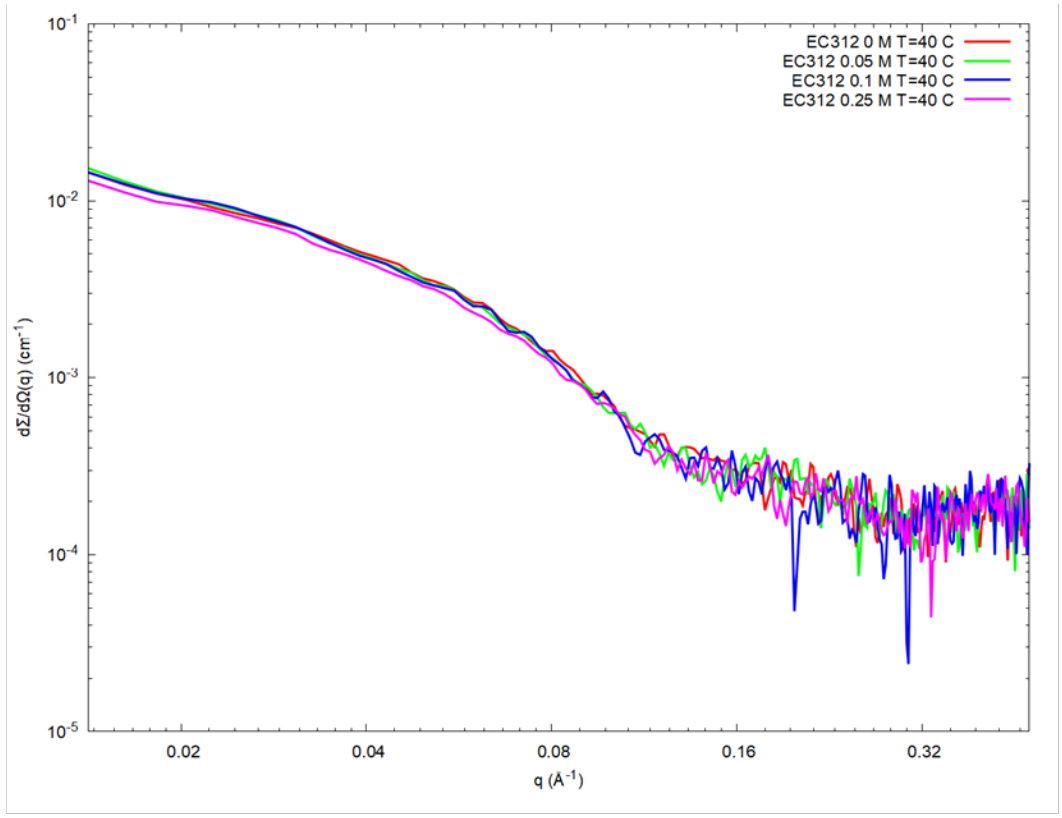


Figure 4.16B : SAXS plots which report ACE2 with sugars at 30°C.



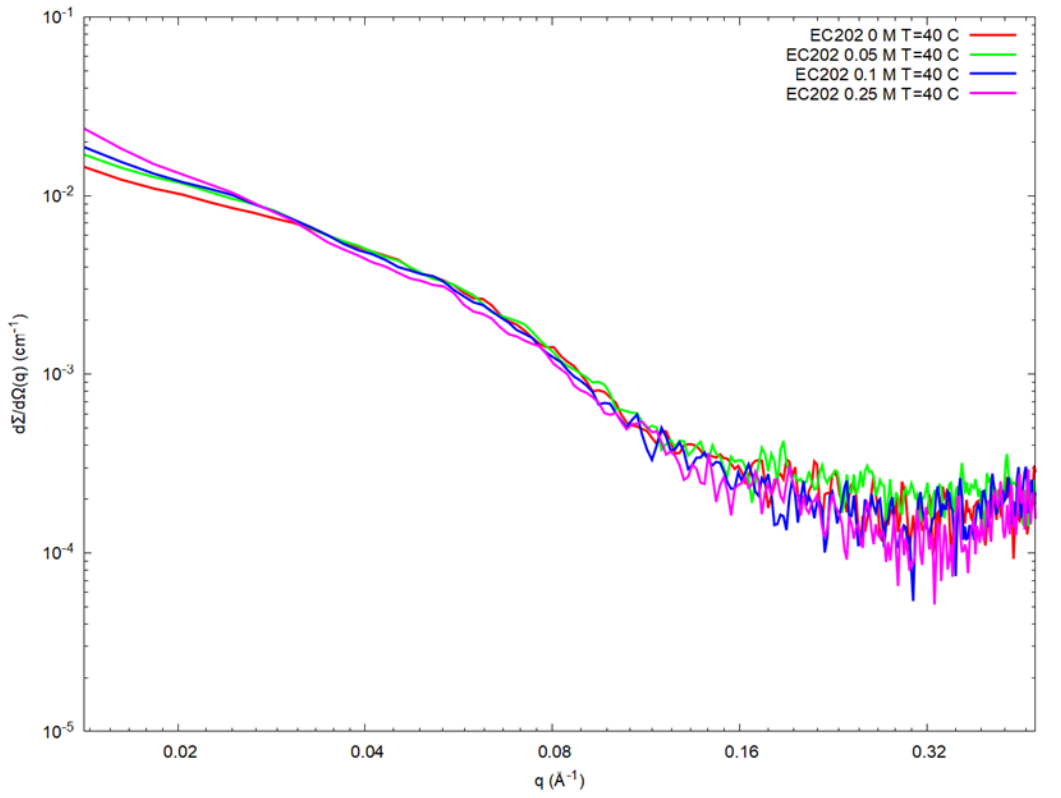


Figure 4.16C : SAXS plots which report ACE2 with sugars at 40°C.

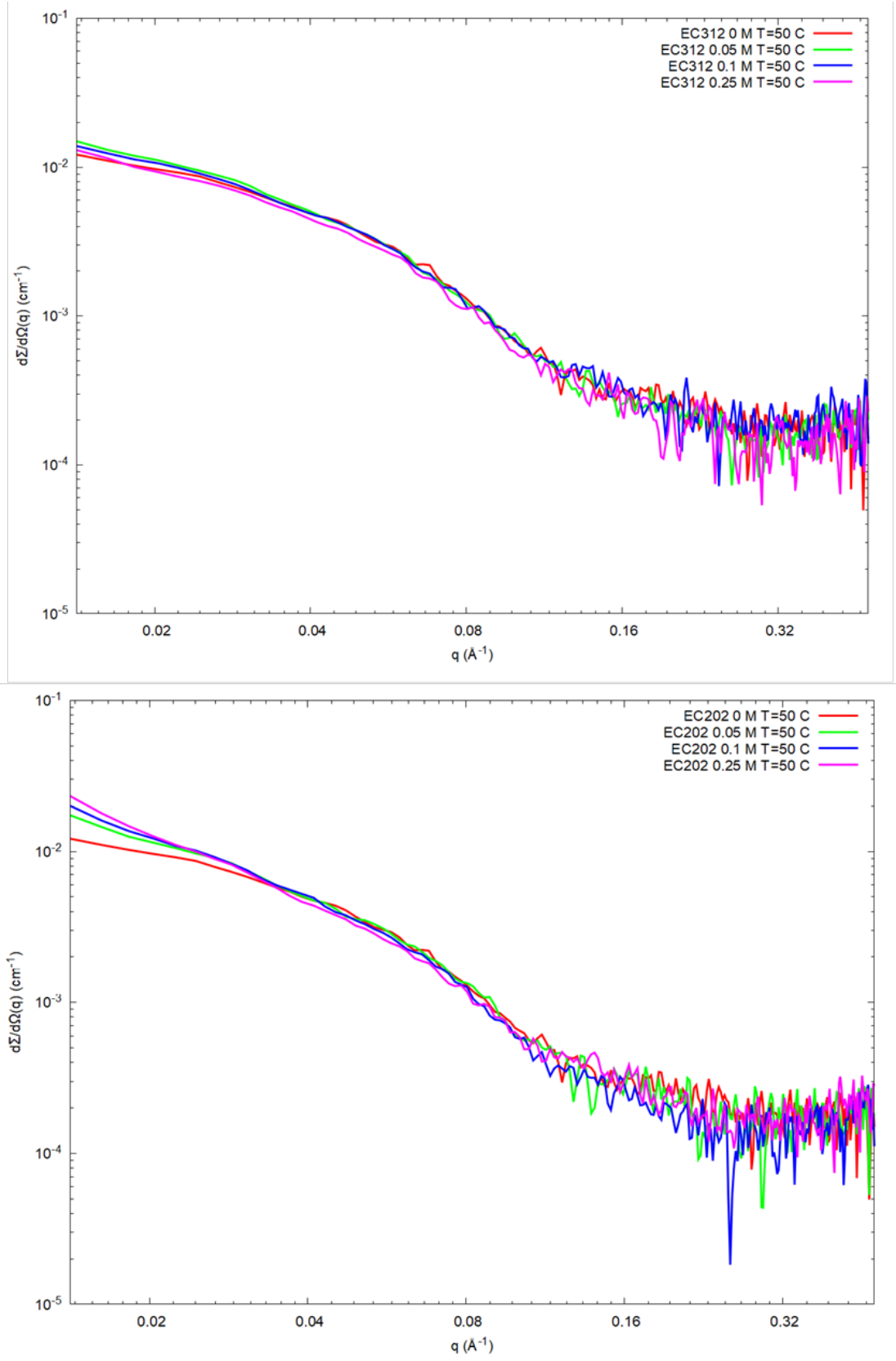


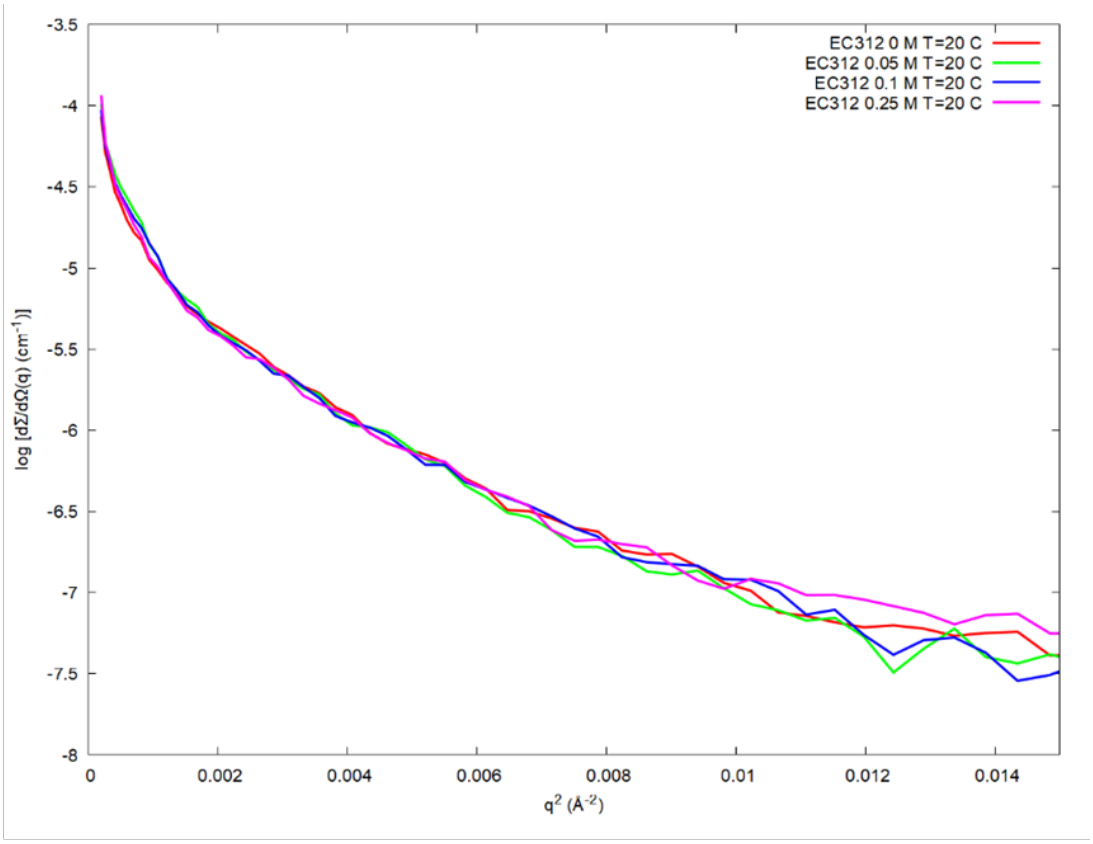
Figure 4.16D : SAXS plots which report ACE2 with sugars at 50°C.

It can be seen that for all the temperatures each sugar has a typical trend. Looking at very low *Q* values, which represent small values of scattering angle, curves obtained for samples with EC-312 at different concentrations are overlapped with the curve of ACE2 without sugar for each temperature. In contrast, for EC-202 curves are not overlapped but, at low *Q*, the SAXS intensity increase with sugar concentration. This could indicate that sugar EC-202 provokes protein aggregation and that the aggregate particles growing up with the sugar concentration.

4.2.2 Guinier plots

The Guinier analysis has been applied to the SAXS curves. To this aim, Guinier plots, which show the logarithm of the intensity as a function of Q^2, are shown.

As in the previous situations, Guinier plot were carried out at a fixed temperature and by changing the sugar concentrations (*Figure 4.17A, B, C, D*).



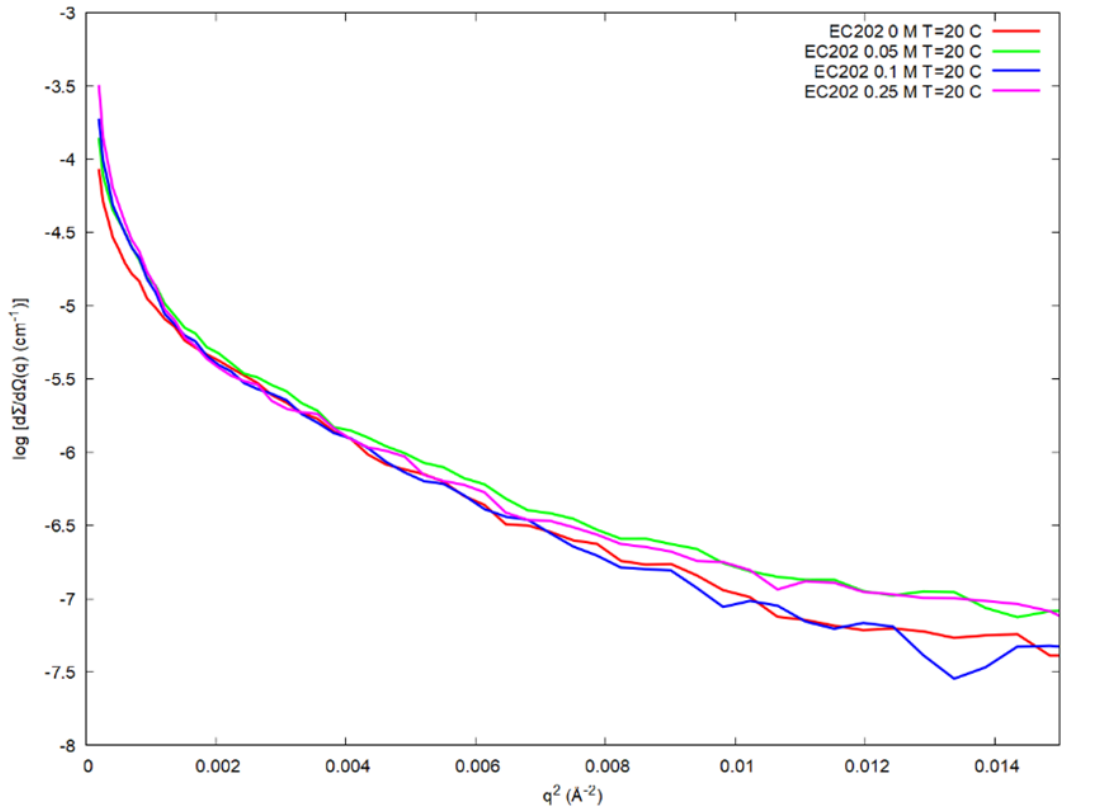
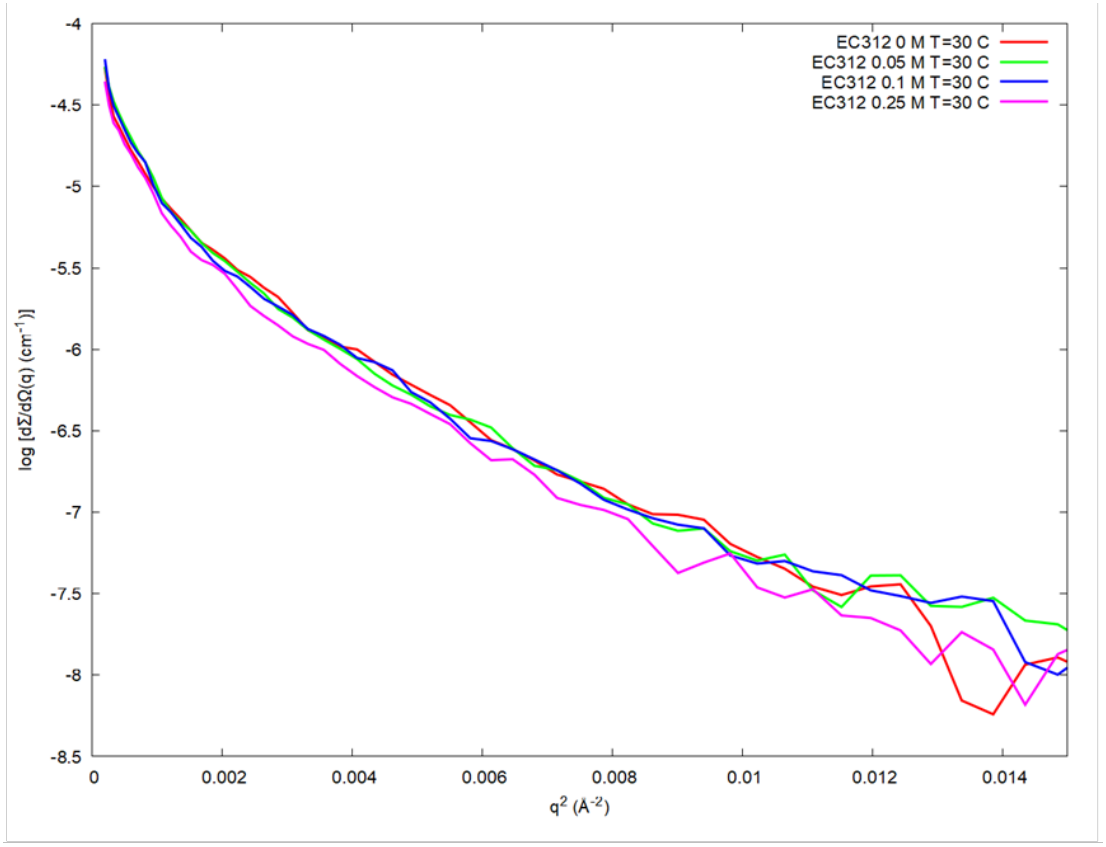


Figure 4.17A : Guinier plots for ACE2 with sugars at 20°C.



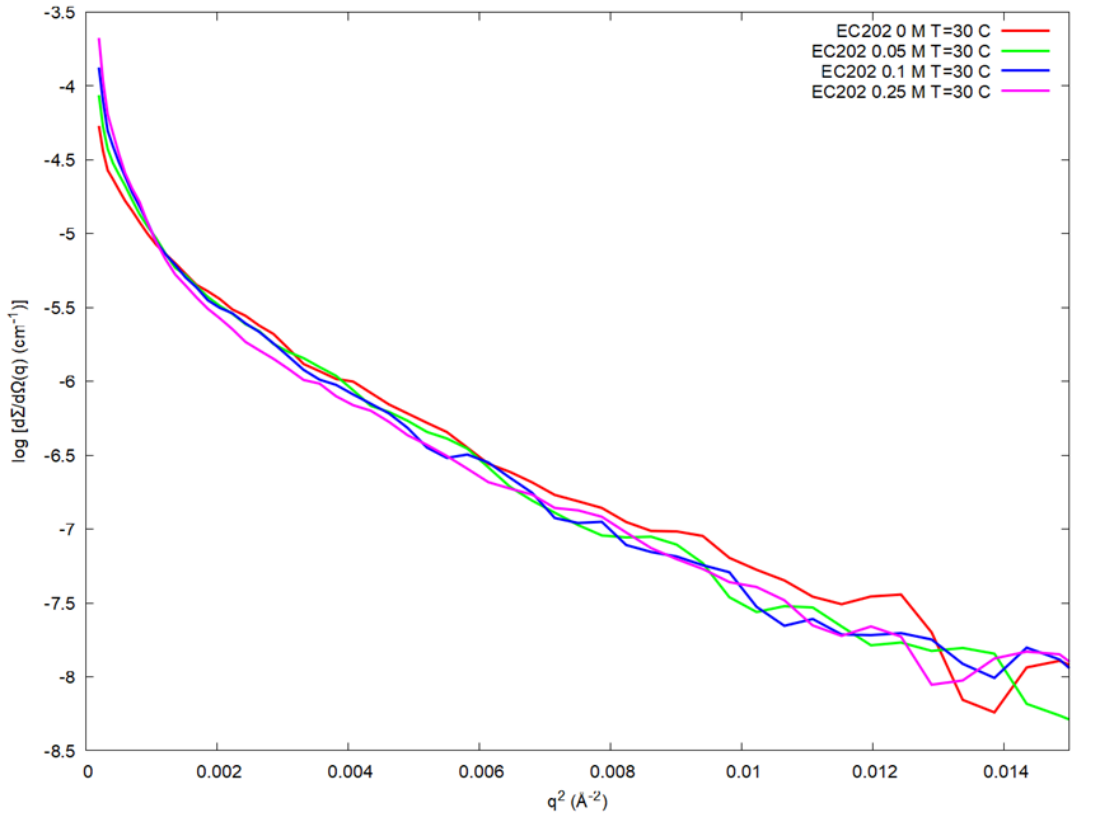


Figure 4.17B : Guinier plots for ACE2 with sugars at 30°C.

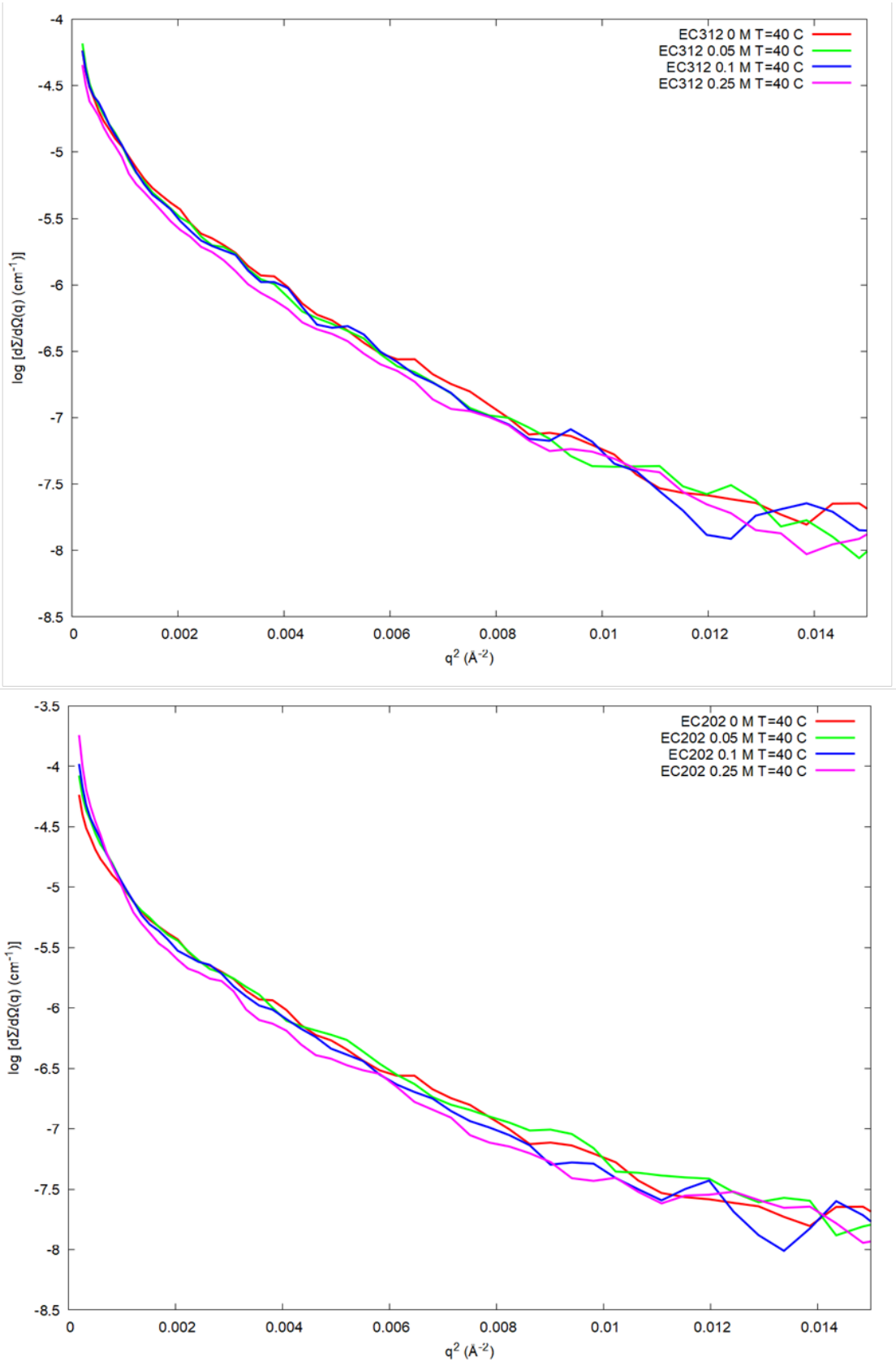
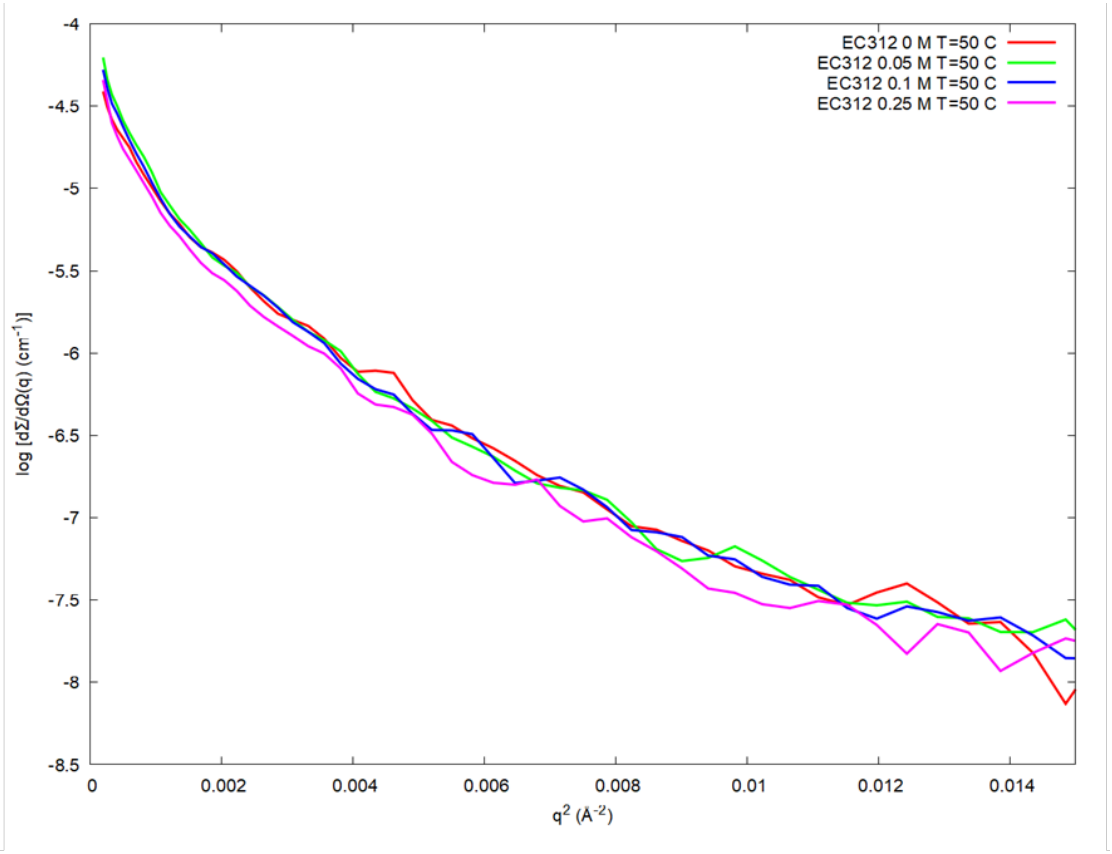


Figure 4.17C : Guinier plots for ACE2 with sugars at 40°C.



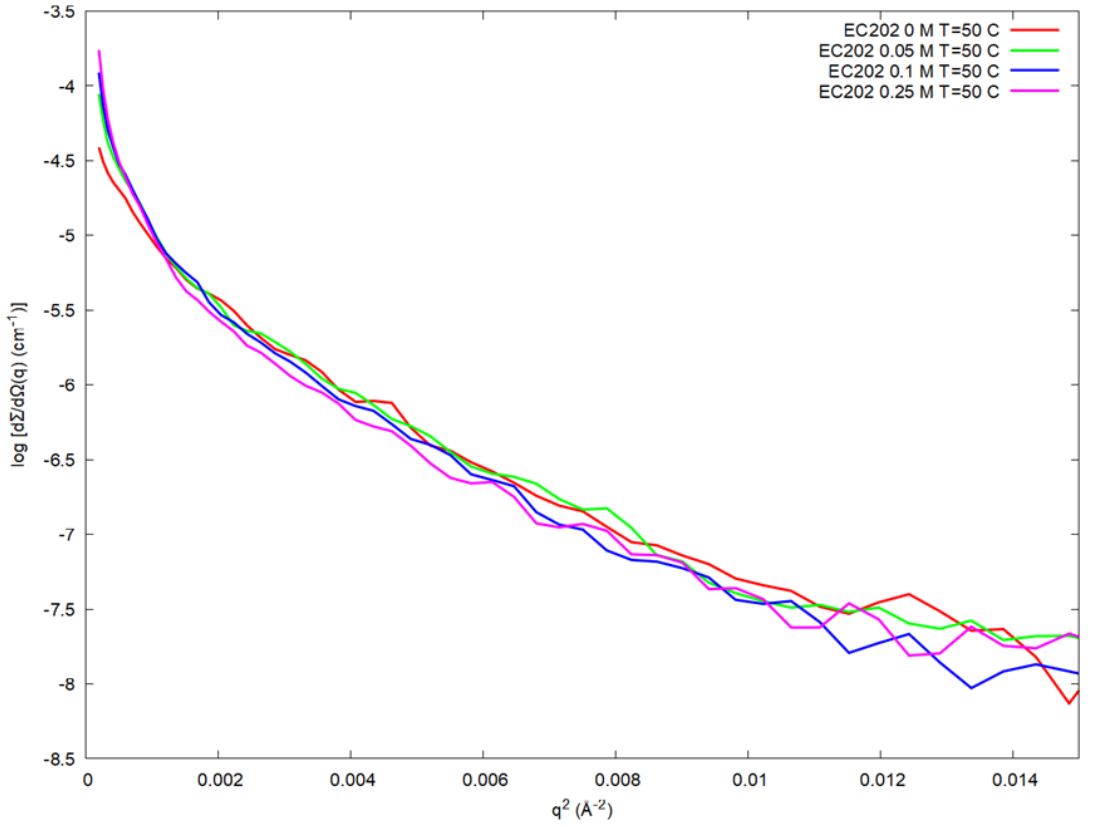


Figure 4.17D : Guinier plots for ACE2 with sugars at 50°C.

Studying the SAXS curves with the Guinier law, information on particle size can be derived. In the case of a globular structure, a linear trend in the Guinier plot, in particular at the lowest values of *Q*, would be detected. In the cases of our samples, each curves rises rapidly upward at low values of Q. The lack of linearity at low values of scattering vector suggests the presence of some aggregates.

4.2.3 Kratky plot

Kratky plots give information about the compactness of particles in solution. This aspect is really important for proteins structure analysis because the compactness degree is related to the conformation state of protein. A protein in its native state or uncompleted unfolded state has totally or partially globular structure. This situation is represented in Kratky plot as a peak. When also there is an unfolded portion, or protein is totally unfolded, there is a plateau or an upward of the curves at high Q values.

This plot reports the intensity multiplied by the square of the scattering vector modulus, $Q^2I(Q)$, as a function of Q.

Also in this case data was plotted in a graph with a fixed temperature and changing the sugar concentrations (*Figure 4.18A, B, C, D*).

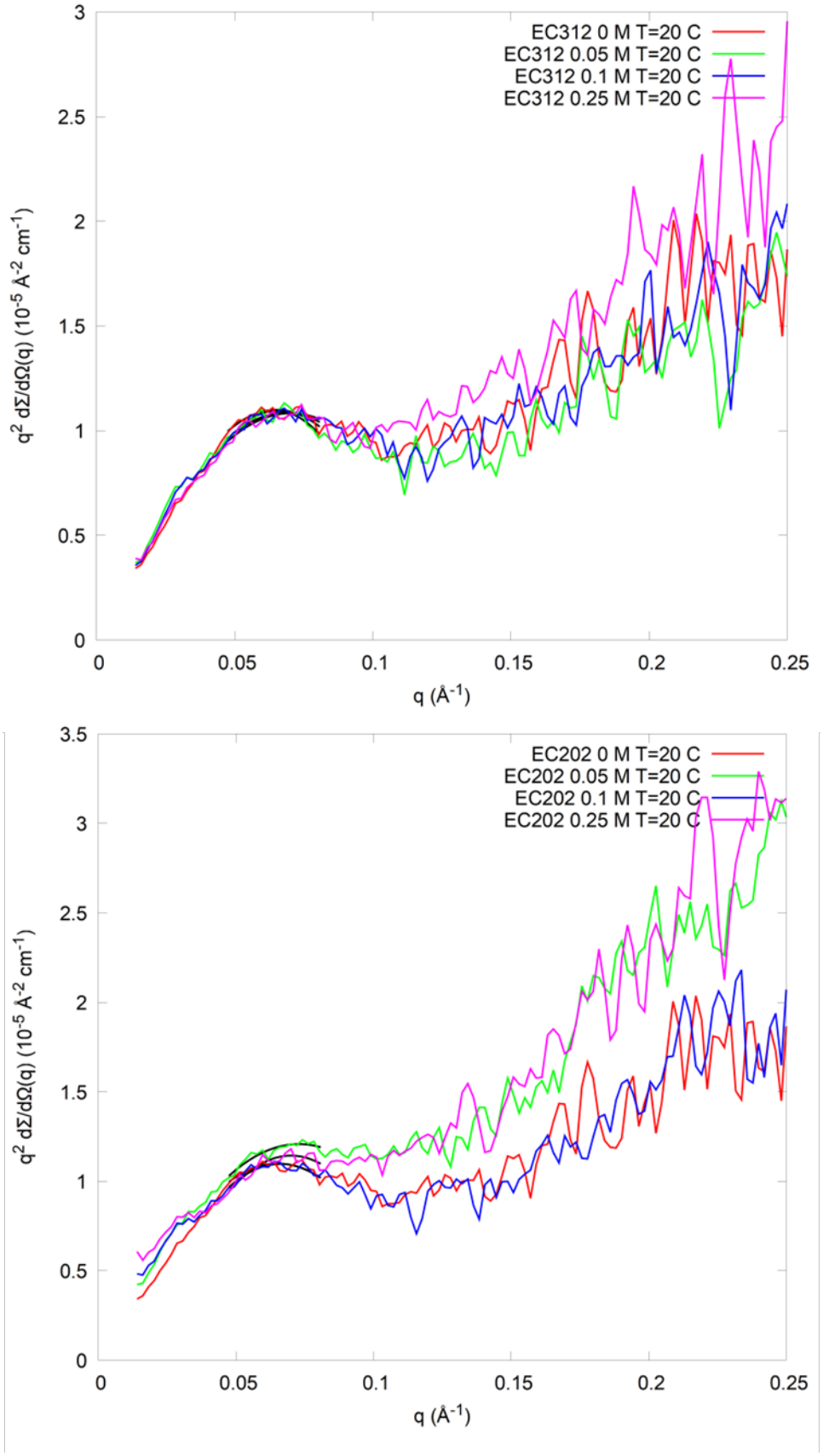


Figure 4.18A : Kratky plots for ACE2 with sugars at 20°C.

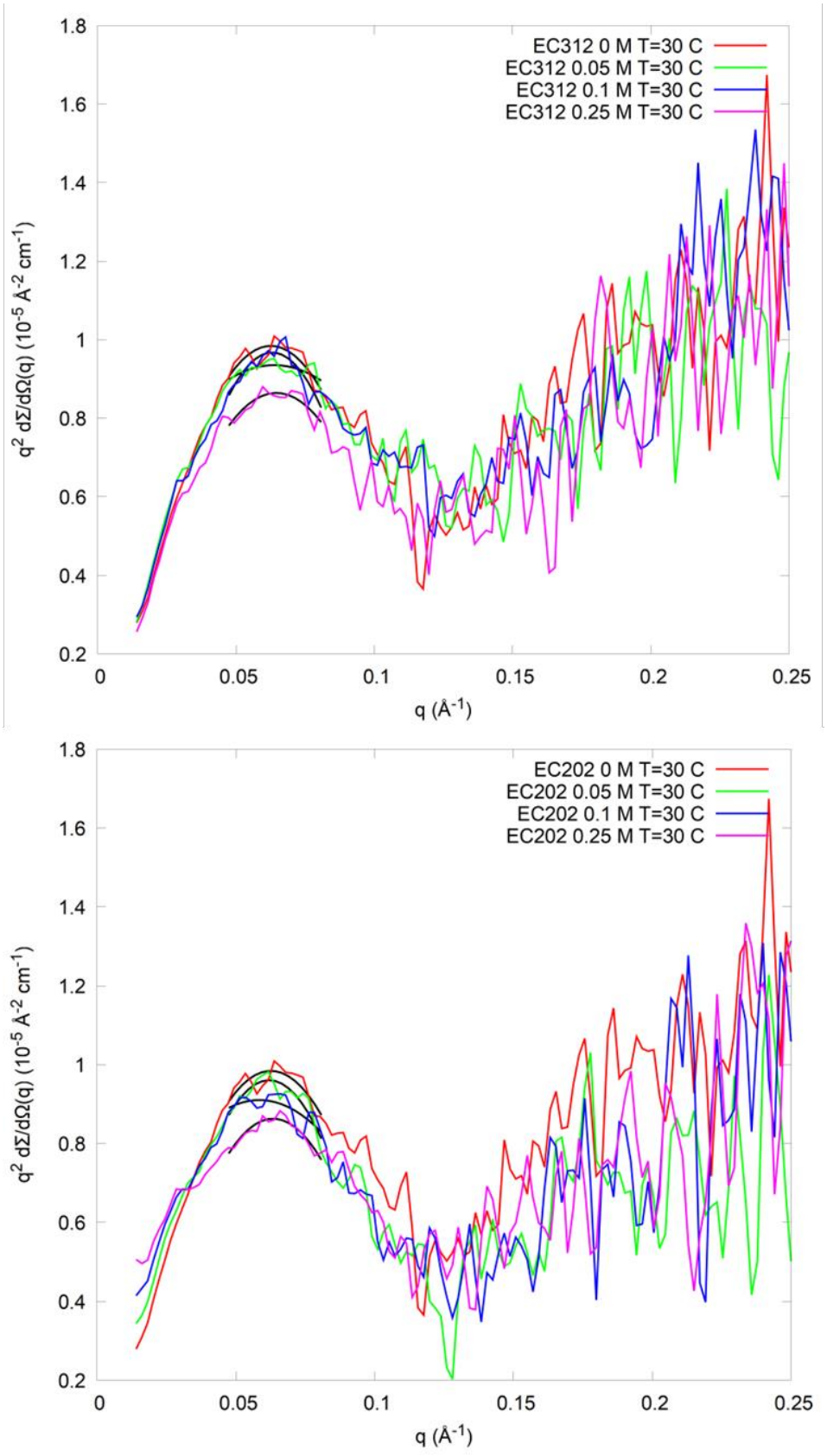
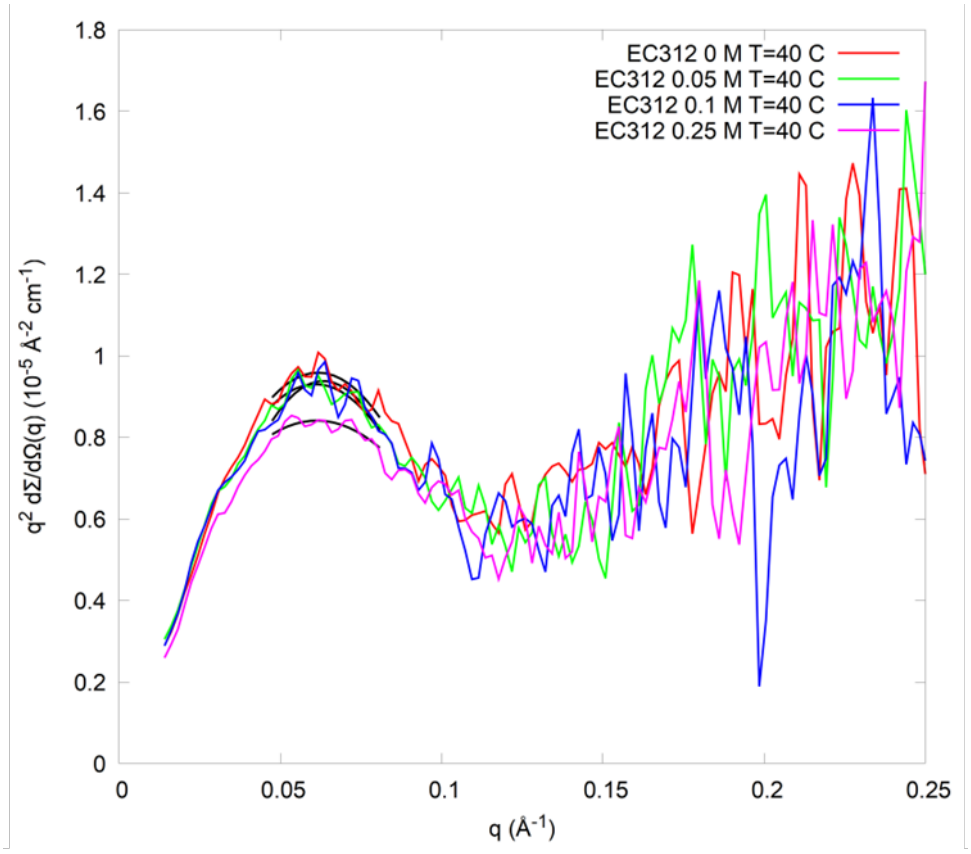


Figure 4.18B : Kratky plots for ACE2 with sugars at 30°C.



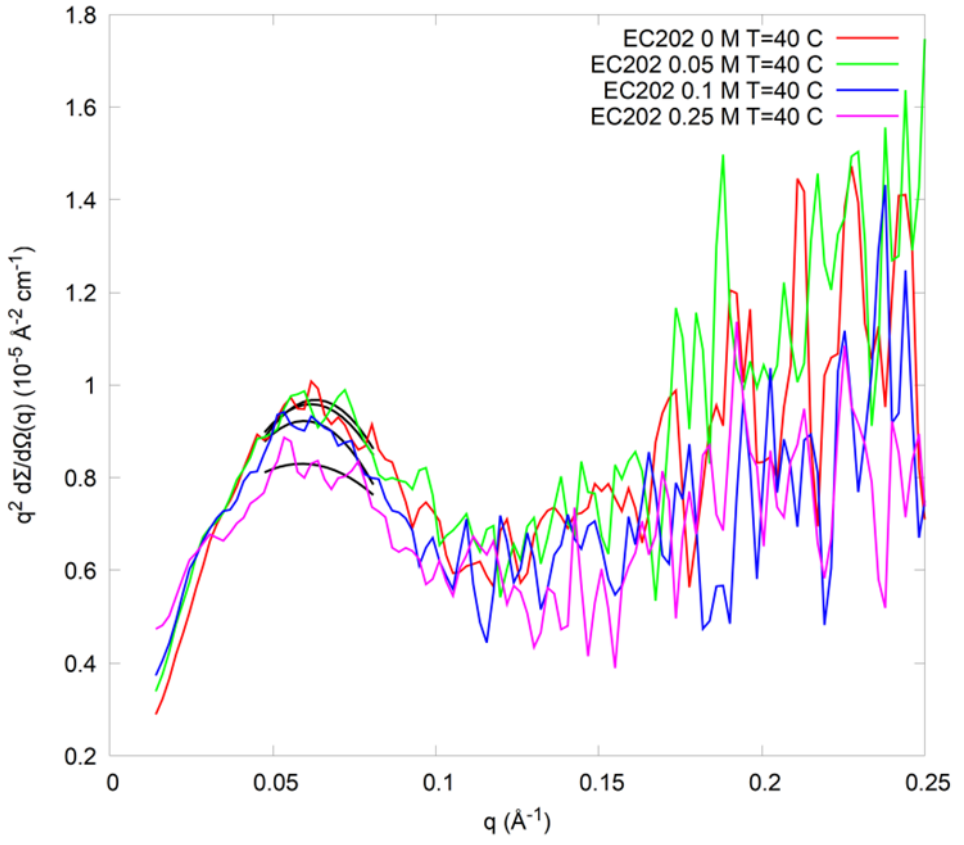
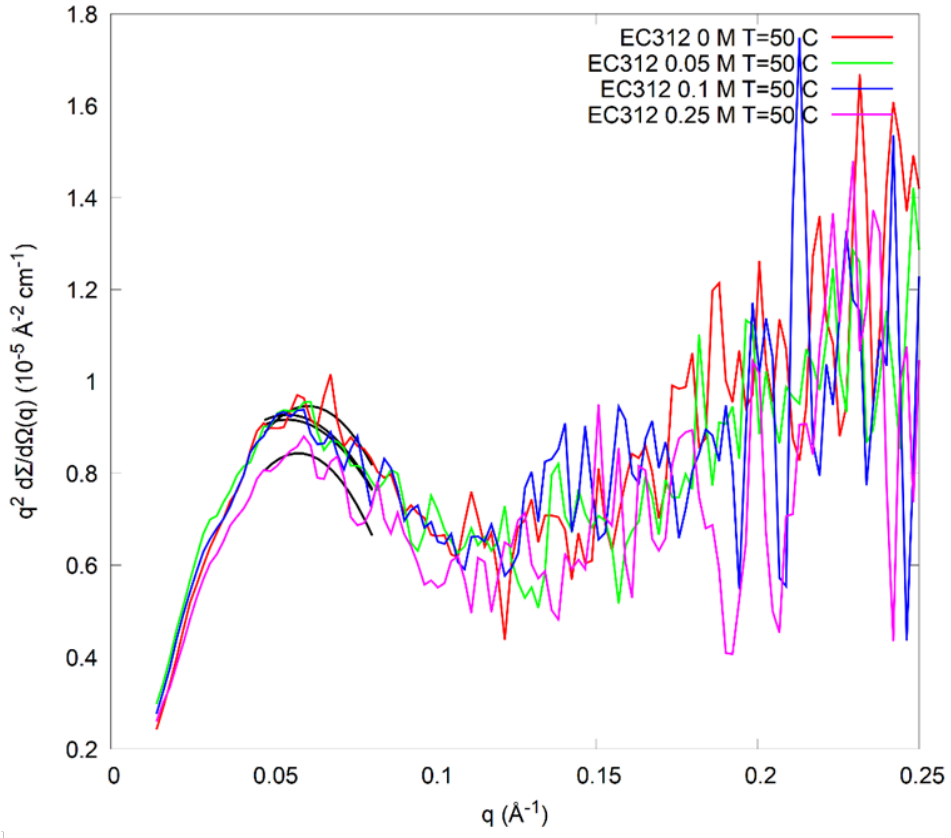


Figure 4.18C : Kratky plots for ACE2 with sugars at 40°C.



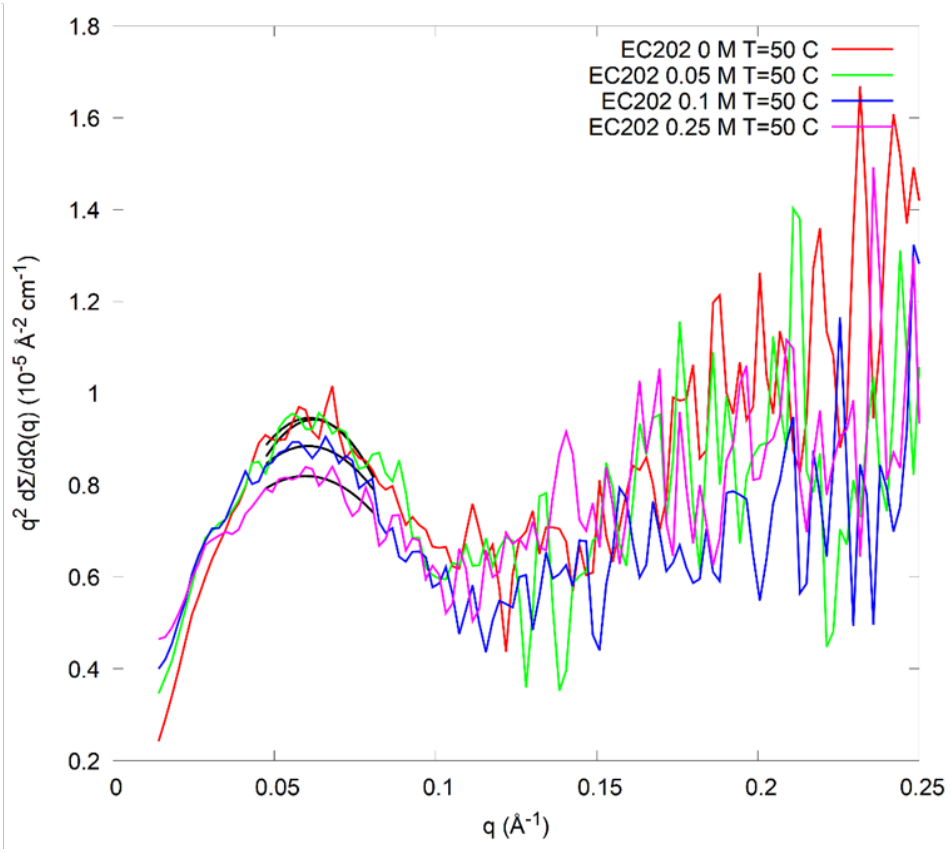


Figure 4.18D : Kratky plots for ACE2 with sugars at 50°C.

In all these situations, it is possible to see a peak at low Q, then curve rises up. This indicates that particles are partially unfolded but maintain a globular part.

4.2.4 Gyration radius

Since a globular region has been detected by the Kratky plots, it was possible to derive the gyration radius just of the folded region by analysing the Kratky plot for each curves. For this purpose we use a formula which consider the maximum Q value of the peak in the Kratky plot (*Eq. 17*).

$$R_g = \frac{1}{3}\frac{2\pi}{Q_{max}} \qquad (17)$$

Q_{max} was obtained performing a fit of the Kratky plot. Looking at the Kratky plots in the previous paragraph, a parabolic fit (black lines) passing through each peaks region has been calculated. We used a script to calculate the parabola parameters. Automatically the script identifies value of Q_{max}. Applying *Eq. 17* we have calculated the gyration radius for each curves, in different conditions of temperature and sugar concentration.

Results are reported in the following tables (*Table 4.1A, B, C, D*), one for each temperature.

Sugar	Sugar concentration (M)	Gyration radius (Å)	Error
no sugar	0	32	0.5
EC-312	0.05	29.8	0.8
EC-312	0.1	30.6	0.6
EC-312	0.25	32	0.4
EC-202	0.05	28.7	0.9
EC-202	0.1	31.6	0.4
EC-202	0.25	29.9	0.7

Table 4.1A : Gyration radius for samples at 20°C.

Sugar	Sugar concentration (M)	Gyration radius (Å)	Error
no sugar	0	33.4	0.5
EC-312	0.05	32.8	0.7
EC-312	0.1	33.2	0.4
EC-312	0.25	33.4	0.4
EC-202	0.05	33.8	0.4
EC-202	0.1	35.8	1.9
EC-202	0.25	33	0.2

Table 4.1B : Gyration radius for samples at 30°C.

Sugar	Sugar concentration (M)	Gyration radius (Å)	Error
no sugar	0	33.9	0.7
EC-312	0.05	34.4	0.7
EC-312	0.1	33.1	0.6
EC-312	0.25	34.1	0.6
EC-202	0.05	33.2	0.7
EC-202	0.1	35.1	0.6
EC-202	0.25	35.5	2.6

Table 4.1C : Gyration radius for samples at 40°C.

Sugar	Sugar concentration (M)	Gyration radius (Å)	Error
no sugar	0	34.4	0.8
EC-312	0.05	38.4	1.8
EC-312	0.1	38.4	2.7
EC-312	0.25	36.1	1.1
EC-202	0.05	33.6	0.5
EC-202	0.1	34.6	0.6
EC-202	0.25	35.1	1.3

Table 4.1D : Gyration radius for samples at 50°C.

The behaviour of the gyration radius is hereafter discussed. Considering the sugar type, in each temperature, it is possible to notice that there are no difference between the effect of EC-312 and EC-202. Now, looking at their concentration, the value of gyration radius does not seem to increase in a linear

way with the sugar concentration. The principal factor which lead to the grow of the gyration radius of the globular part of particles appears to be the temperature. In fact the value of Rg increase proportionally to the temperature. It is easy to notice that by calculating the average of gyration radius for each temperature. At temperature of 20°C the average gyration radius, with relative error, is 30.6 ± 0.6, at 30°C is 33.6 ± 0.6, at 40°C is 34.2 ± 0.9 and at 50°C is 35.8 ± 1.2. This trend might indicates that there is an aggregation of globular portion of particles which occurs with temperature increasing.

Chapter V

DISCUSSION

In the previous chapter we showed that it is possible to derive information about the stabilize effect of modified sugars on proteins from DLS and SAXS data obtained by an approach which combines structural parameters and temperature dependence of proteins in solution. We worked on DLS data starting from correlograms, from which intensity weighted size distributions were fitted. Then intensity weighted size distributions were converted in volume fraction distributions and these were analysed using a sum of Gaussians. These Gaussians allow to identify the population of the protein states. For myoglobin, we have considered monomers, dimers and unfold monomer. For insulin, monomers, dimers, tetramers, hexamers and larger structures have been taken into account. SAXS data were analysed initially in qualitative terms, by plotting all the curves to compare ACE2, in solution with the sugars, with the protein in native conditions. Then, in order to derive the protein size, data were analysed with the Guinier law. We also visualized data in the Kratky plot, which is particularly useful to identify globular proteins states.

5.1 DLS data discussion

Based on DLS results, it is possible to identify which sugars are more useful for proteins stabilization. For this purpose, we consider the plots of fraction of monomers distributed in the different protein species calculated for each temperature. First of all, it is possible to notice that results for insulin are more defined than the ones of myoglobin.

Definitely, the worst sugar for protein stabilization is EC-202. In the case of myoglobin, results do not show a specific trend and there are larger structures as well as monomer and dimer for both protein concentration. The effect of this sugar is more evident when added to insulin: the protein immediately forms very large aggregates and precipitated. Moreover this happened also with the same sugar at 0.1 M (shown in *Appendix*). Also EC-101 leads protein aggregation but not as strong as EC-202. In fact insulin, with this sugar, aggregates and precipitates just in a concentration of 0.25 M. The reduced aggregation effect of EC-101, compared to EC-202, is also visible for myoglobin. The two concentrations of myoglobin show different results. Indeed, the sample at 2 g/L shows the presence of dimer in addition to monomer, and in the sample at 4 g/L there are also larger aggregates, which are absent for the other concentration.

Proteins associated with the other two sugars have shown a completely different behaviour. Samples with EC-212 and EC-312 have similar trends for each protein, but EC-312 results the best sugar to stabilize proteins against aggregation. Myoglobin, in both concentration, is mainly in monomeric conformation, with a little amount of dimers and unfolded monomers, especially for the samples with EC-212 and the one with EC-312 and myoglobin at 4 g/L. In myoglobin at 2 g/L, sample in presence of EC-312 are dimers and other larger conformations are almost totally absent. The effects of these sugars are more evident for insulin, in fact in addition to monomer and dimer conformation there is a little amount of tetramer and hexamer when there is EC-212. Insulin with EC-312 is completely in monomeric and dimeric, states without the presence of other bigger conformations.

5.2 SAXS data discussion

Based on SAXS results, it is possible to make observations about sugars role on proteins stabilization and also about the protein conformation. For the first purpose, we consider the curve behaviour at low Q values. With sugar EC-312, curves of the three concentrations, compared with the one of ACE2 without sugar, are overlapped and it is not possible to notice any differences. On the contrary, samples which contain EC-202 present differences between them and with respect to the sample without sugar, and also between the three

concentrations. At low Q, the intensity of SAXS curves linearly increases with sugar concentration increases. This indicates protein aggregation. The behaviour of ACE2 protein in presence of EC-312 does not indicate that protein aggregation. Comparing both sugars, it emerges that EC-312 is the best stabilizer sugar because it maintains proteins in their native state. On the contrary, EC-202 leads ACE2 aggregation and it is not useful to preserve proteins. These results confirm what we observed with DLS.

SAXS measurements give also information about protein size and conformation, by the analysis with Guinier law and by reporting data in the form Kratky plots, respectively. In Chapter V, these plots have been described. In particular, we saw that Kratky plot presents a peak and then the curve goes upward. Hence, this qualitative analysis reveals that ACE2 protein is formed by a globular region and with an unstructured region. In addition, we calculated the apparent gyration radius of the globular region for the condition at 20°C. It was found as large as 31 Å. Since we have seen that proteins are aggregated, we have checked whether the value of the apparent gyration radius is due to the size of the globular domain of a single protein or it is the size of protein aggregates. Knowing the molecular weight of ACE2 fraction (35 kDa), the gyration radius of the whole protein, assuming that it has a globular shape, was calculated and resulted about 21 Å. The hypothetical value of gyration radius

of entire protein is less than the value of the protein which has a globular and an unfolded part. This is indicative of an aggregation. A qualitative analysis may suggests that proteins make contacts with each other, probably thanks to the globular regions and the unfolded components are around the compact centre. Folded domains are in the centre of a large particles and surrounded by the unfolded chain which are facing outward.

Chapter VI

CONCLUSION

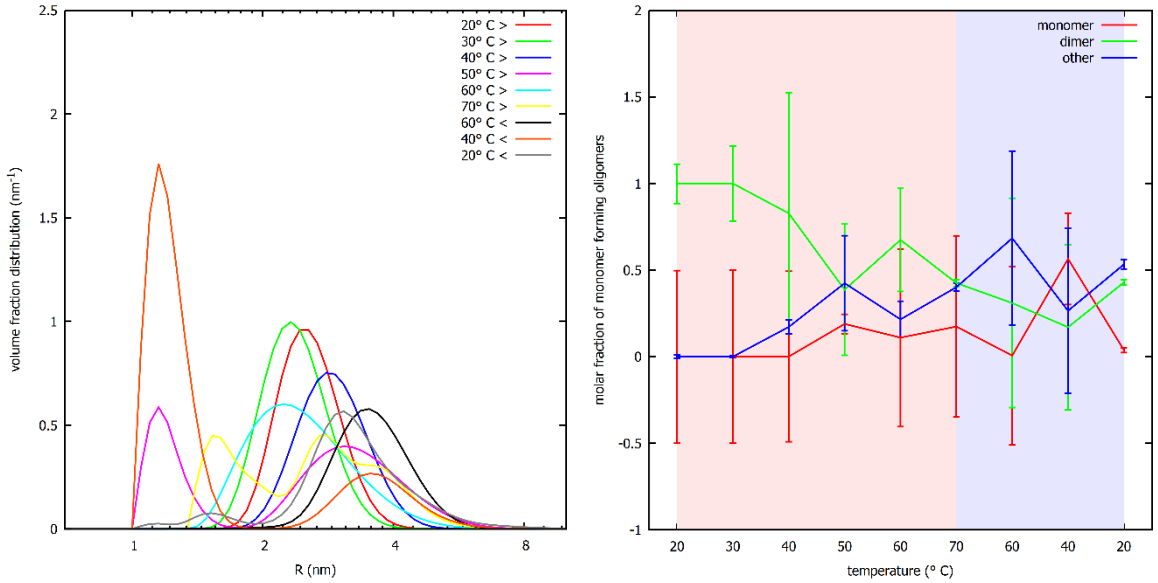
In this thesis we studied the behaviour of models proteins in solution with cosolvents, which are molecular agents with the purpose of stabilize proteins against denaturation. In particular we used some modified sugars as cosolvents synthetized by ExtremoChem. We had five sugars with different characteristics, EC- 212, EC-311 and EC-312 are not electrically charged, while EC-101 and EC-202 have ionic characteristics. DLS and SAXS investigation were performed on samples of different protein and sugar concentrations. It emerged that EC-212 and EC-312 are the best sugars to stabilize proteins because proteins' behaviour in their presence is more similar to the protein without cosolvent. By contrast, samples containing EC-101 and EC-202 contain a greater number of larger particles, hence they lead protein aggregation. Considering the characteristics of each sugar, it is possible to notice that only the charged sugars lead aggregation, while the uncharged sugars preserve the native protein structure. The ionic state of these molecules seem to be the main factor for the protein maintenance. In conclusion, EC-311 should be considered a good cosolvent against protein denaturation.

The aim of this study was to understand the protein's behaviour in solution with cosolvents and to identify the best cosolvent useful for protein

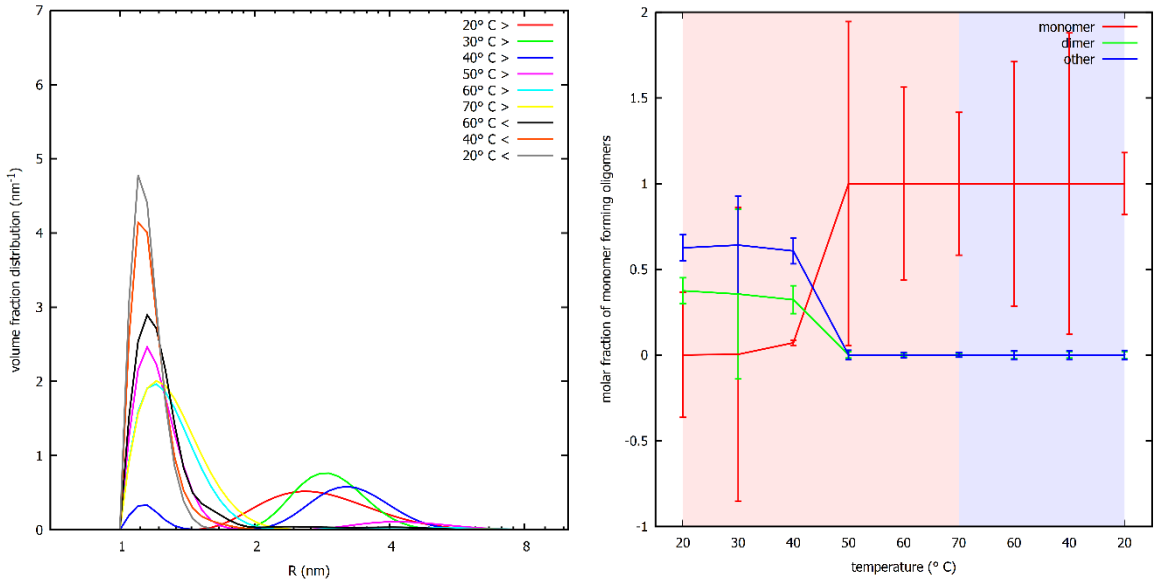
preservation. This aspect is really important because some proteins, in particular the ones which have pharmaceutical importance, during transport and storage should be maintain to preserve their structure, hence their function. They need continuously refrigeration and this process is expensive. Cosolvents can improve this situation and to make proteins conservation easier. Based on our results, sugars EC-212 and EC-312 have proven to be good agents to preserve proteins. They could be used in addition to a protein solution, for example in a pharmaceutical preparation. These cosolvents would make it easier and less expensive protein storage and transportation.

Appendix

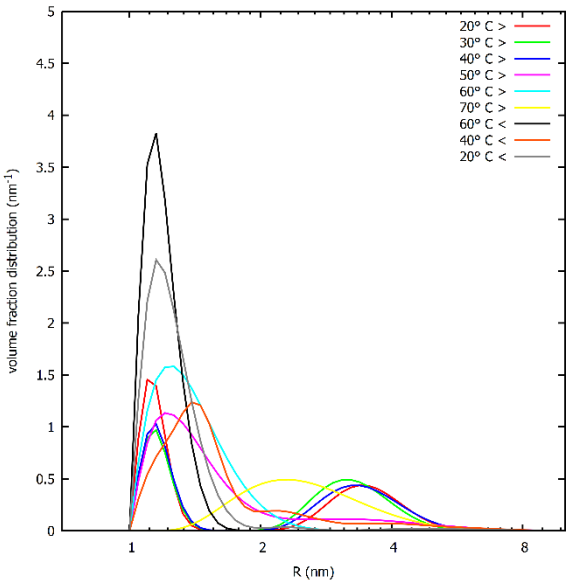
Myoglobin 2 g/L + EC-101 0.1 M

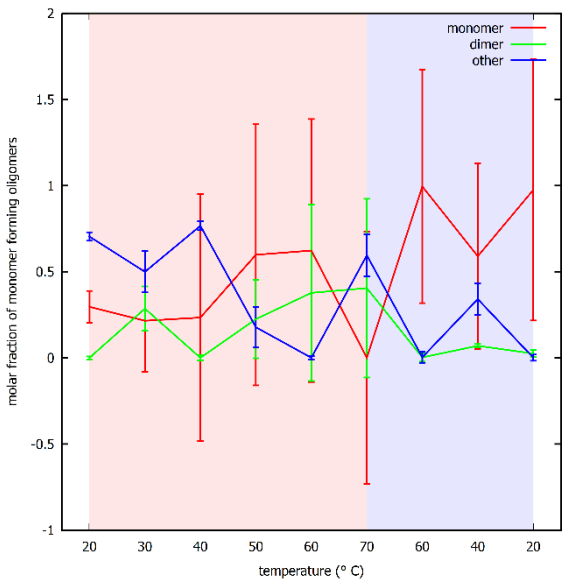


Myoglobin 2 g/L + EC-202 0.1 M

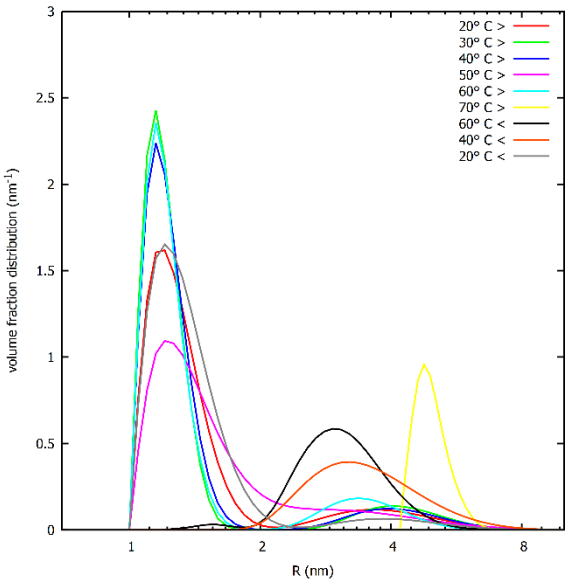


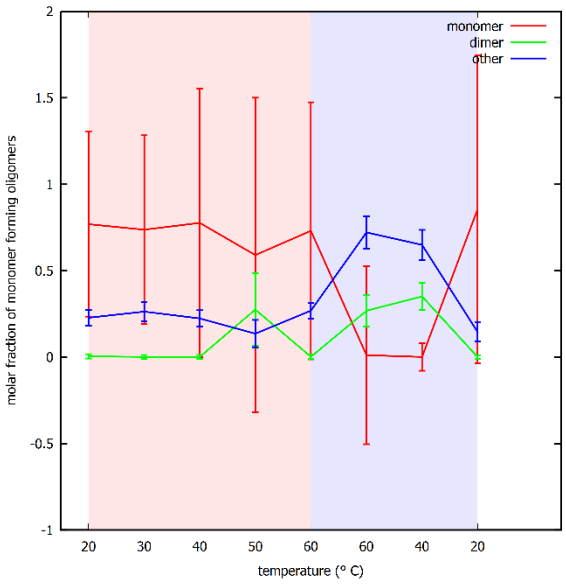
Myoglobin 2 g/L + EC-202 0.05 M



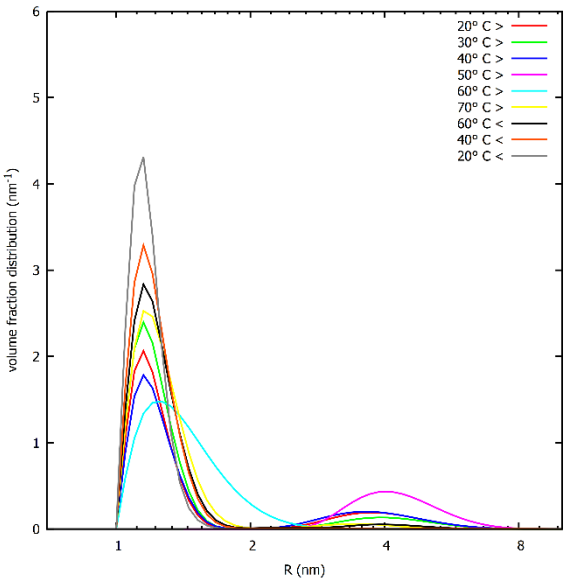


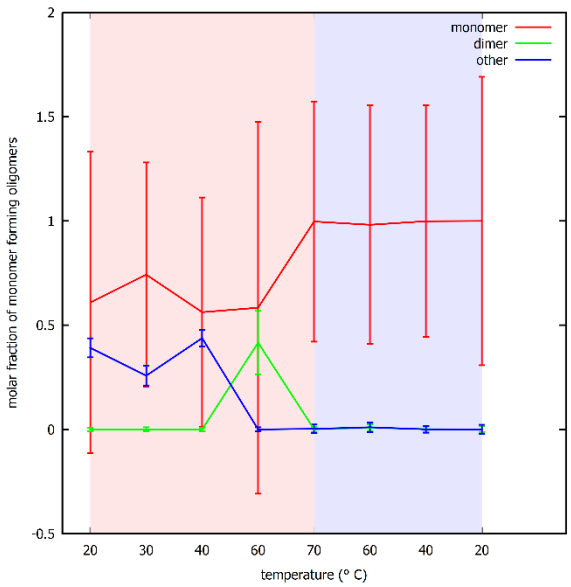
Myoglobin 2 g/L + EC-212 0.1 M



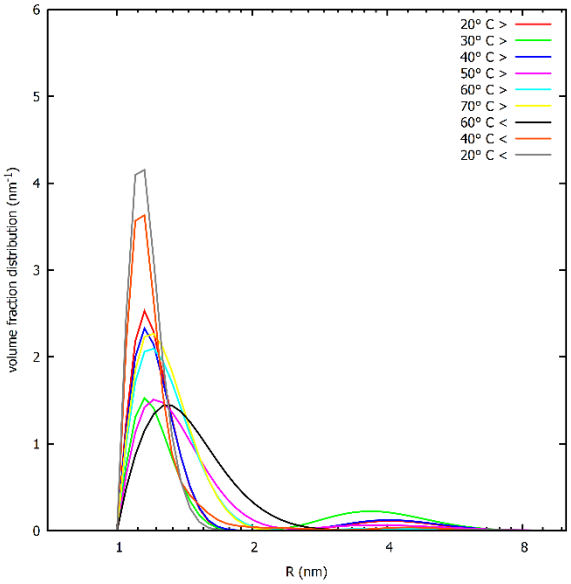


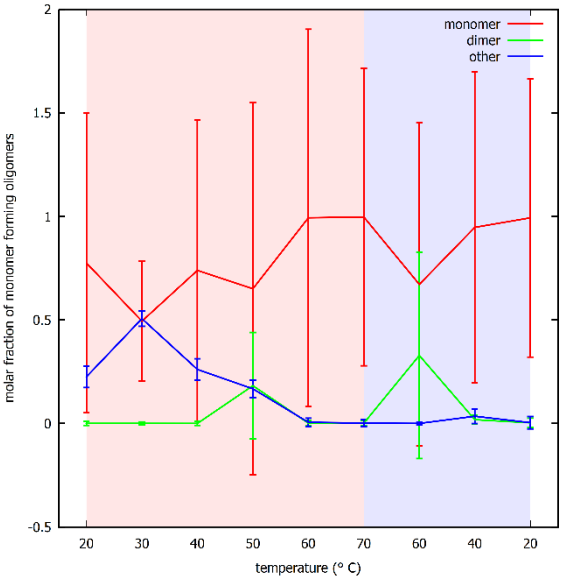
Myoglobin 2 g/L + EC-212 0.05 M



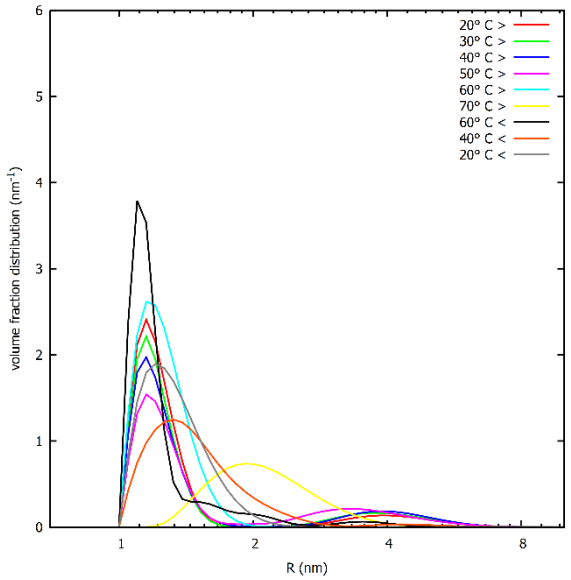


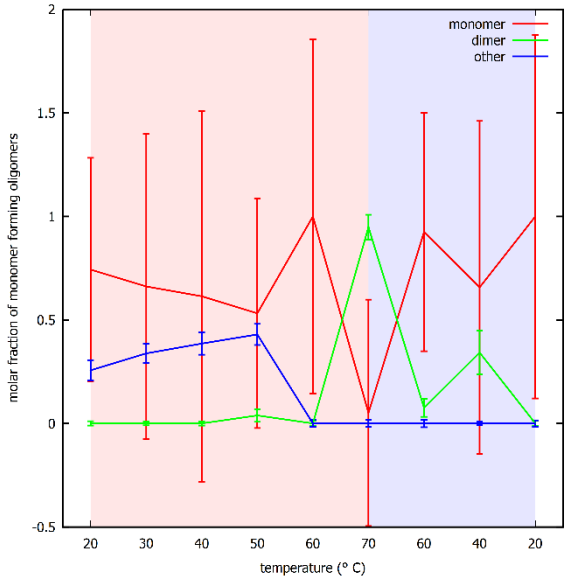
Myoglobin 2 g/L + EC-312 0.1 M



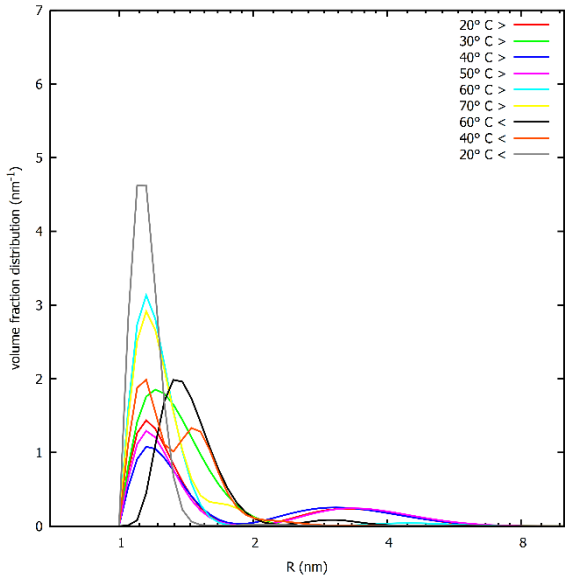


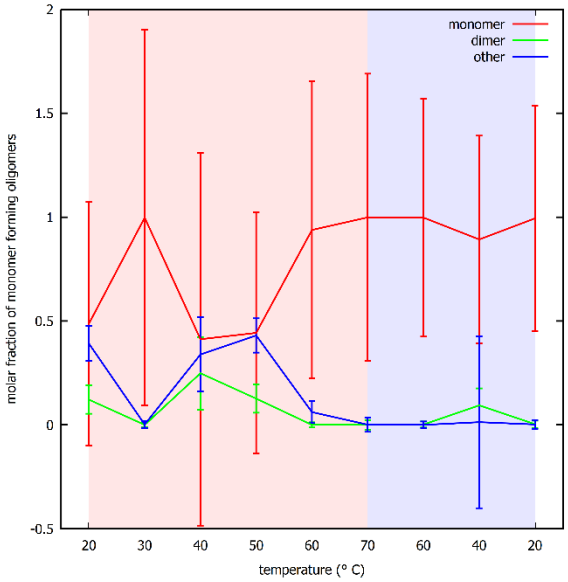
Myoglobin 2 g/L + EC-312 0.05 M



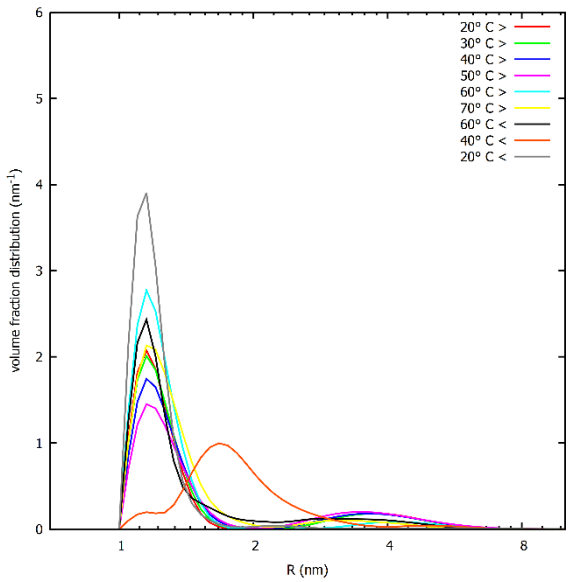


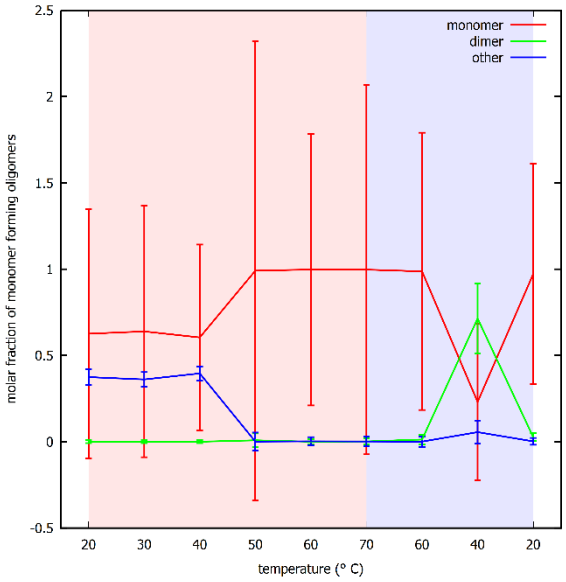
Myoglobin 4 g/L + EC-101 0.05 M



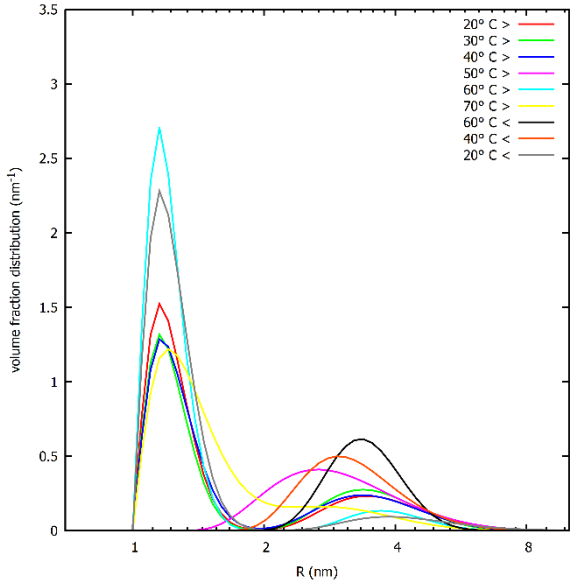


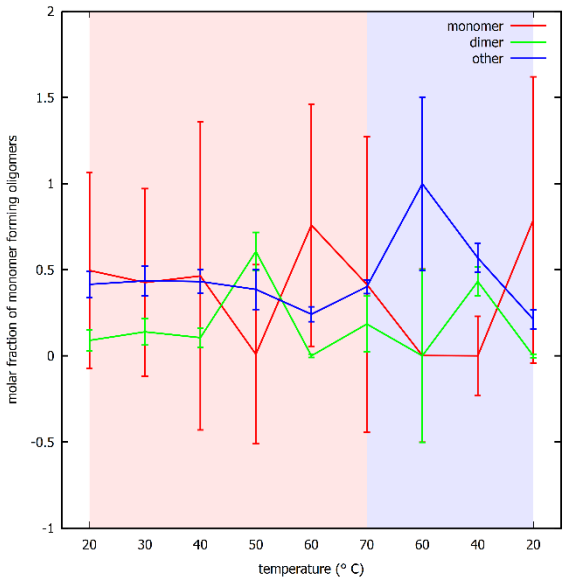
Myoglobin 4 g/L + EC-202 0.1 M



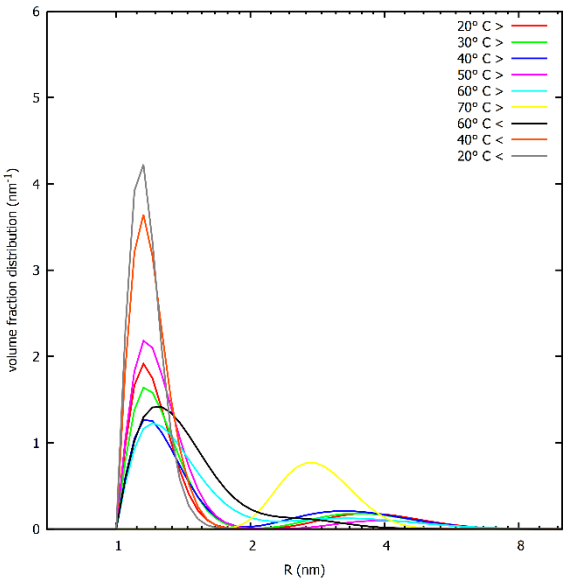


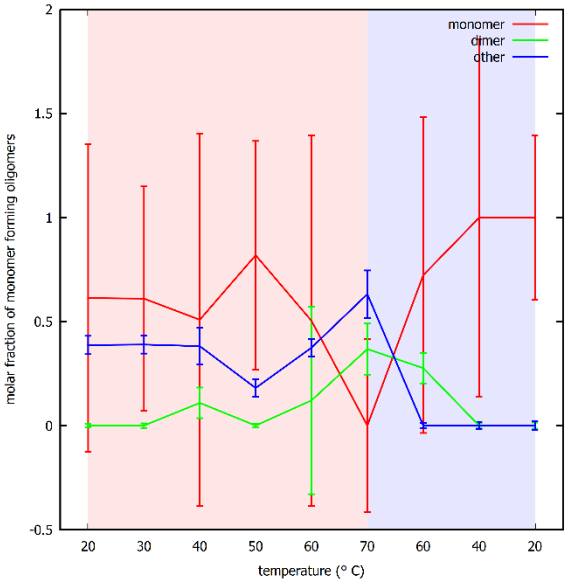
Myoglobin 4 g/L + EC-202 0.05 M



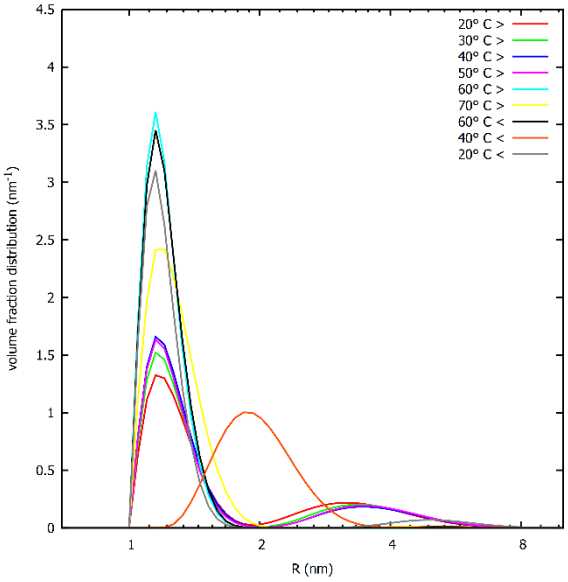


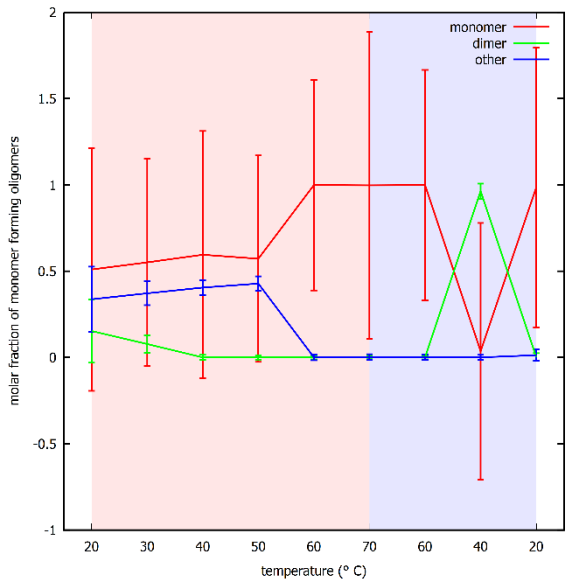
Myoglobin 4 g/L + EC-212 0.1 M



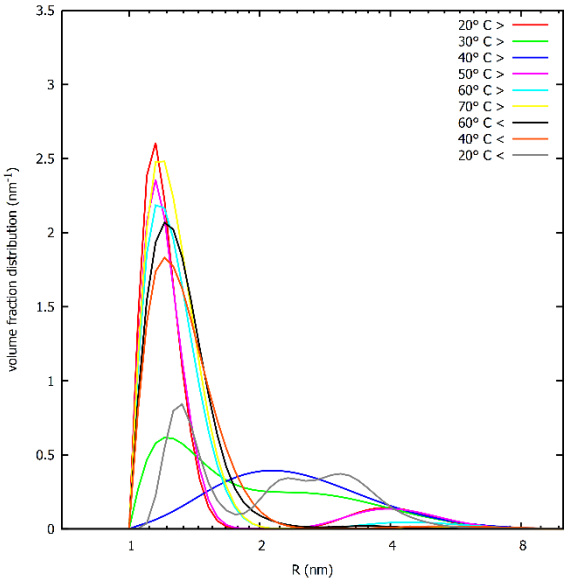


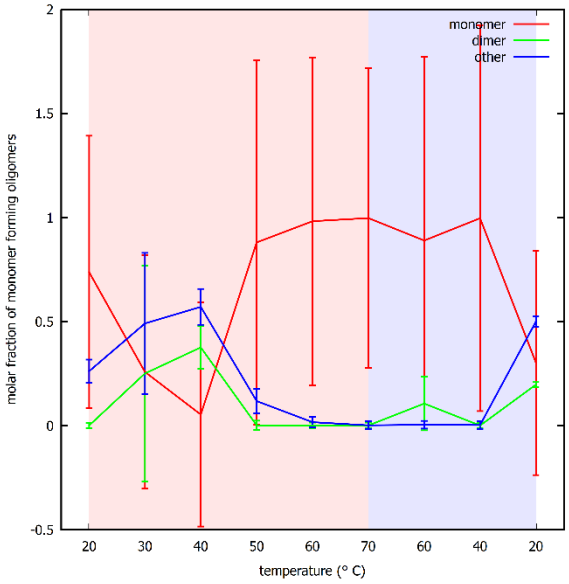
Myoglobin 4 g/L + EC-212 0.05 M



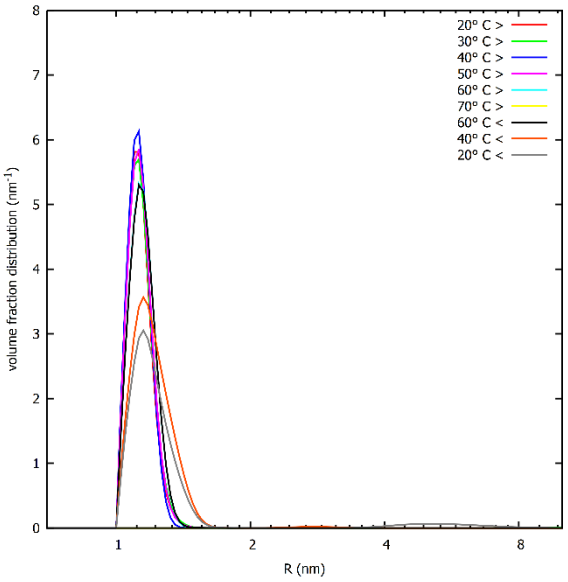


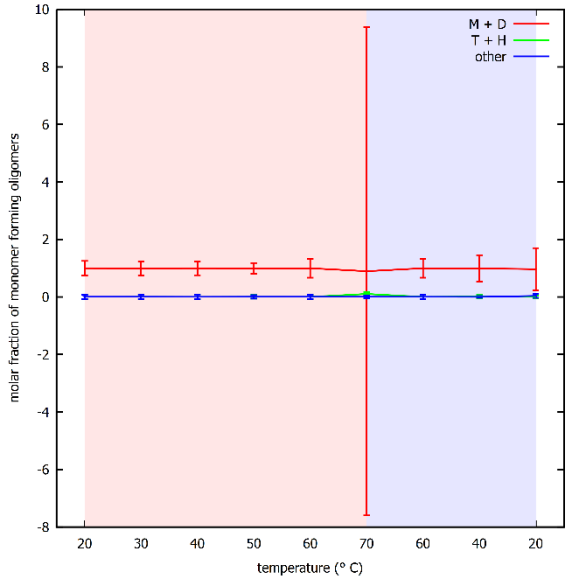
Myoglobin 4 g/L + EC-312 0.05 M



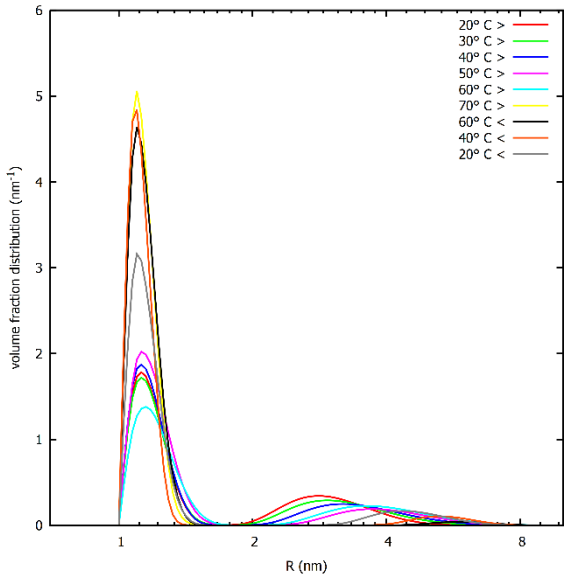


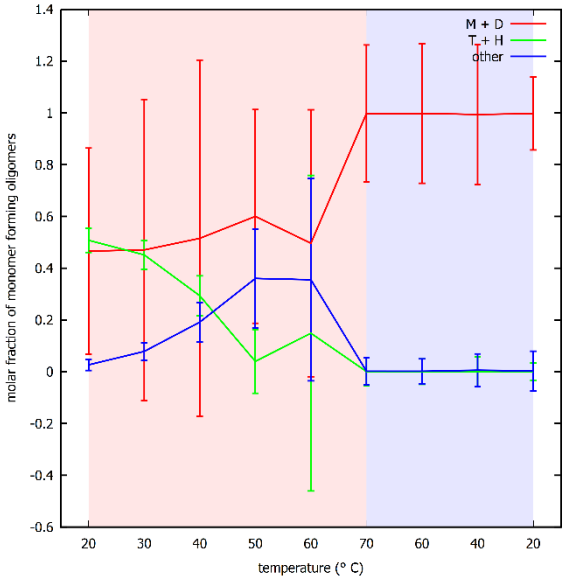
Insulin 2 g/L + EC-101 0.1 M



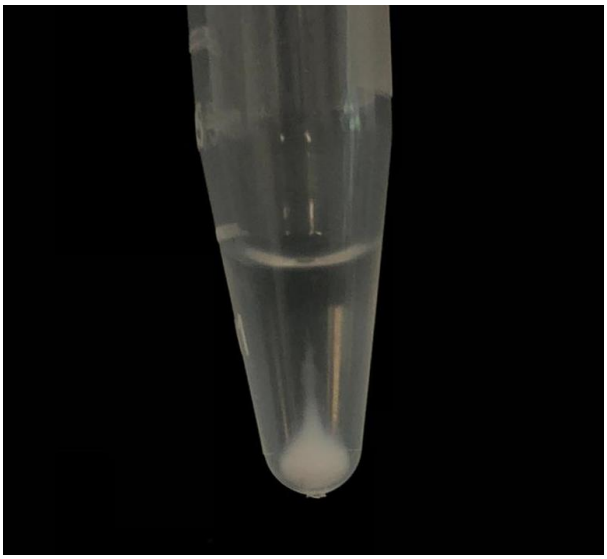


Insulin 2 g/L + EC-101 0.05 M

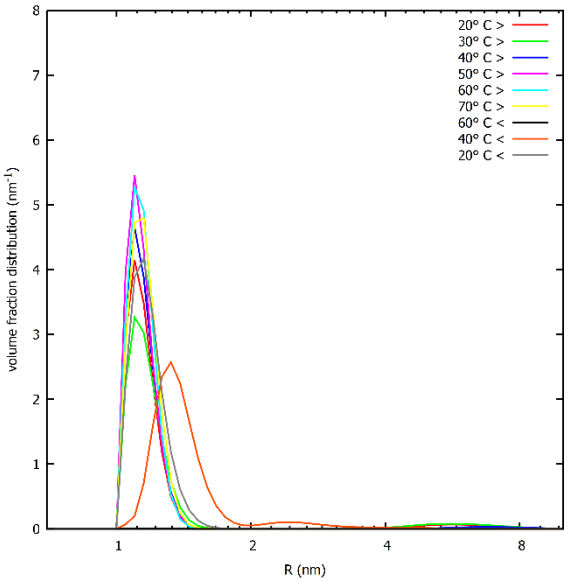


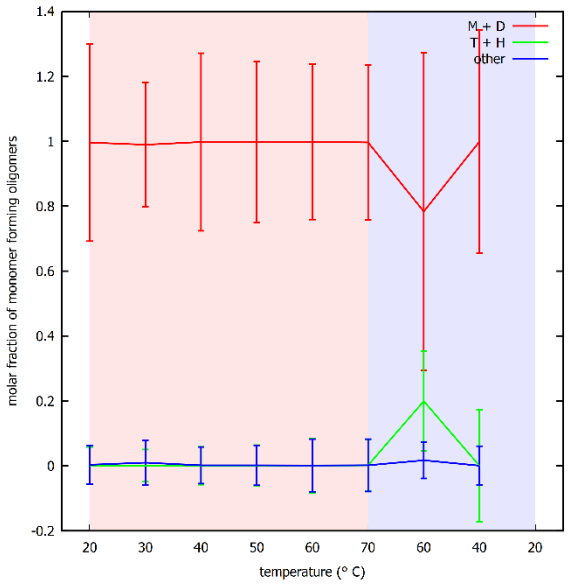


Insulin 2 g/L + EC-202 0.1 M

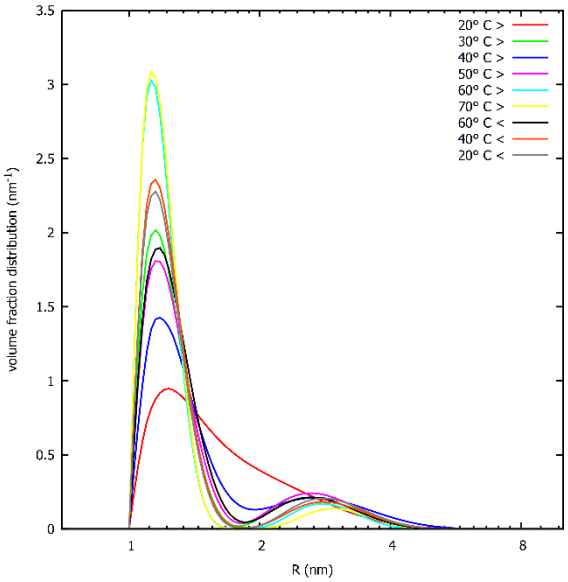


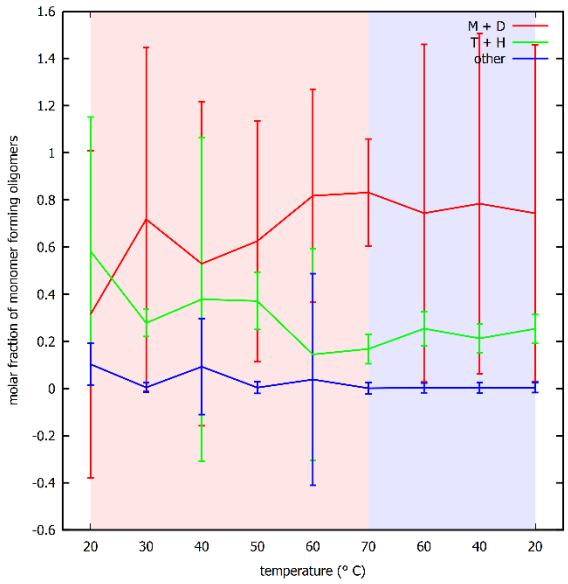
Insulin 2 g/L + EC-202 0.05 M



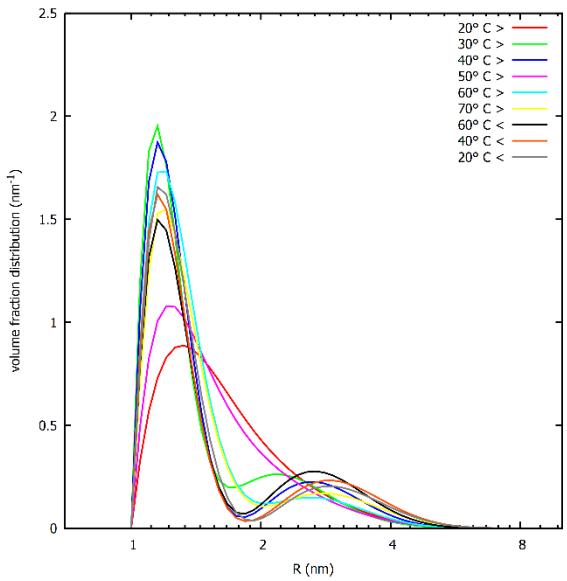


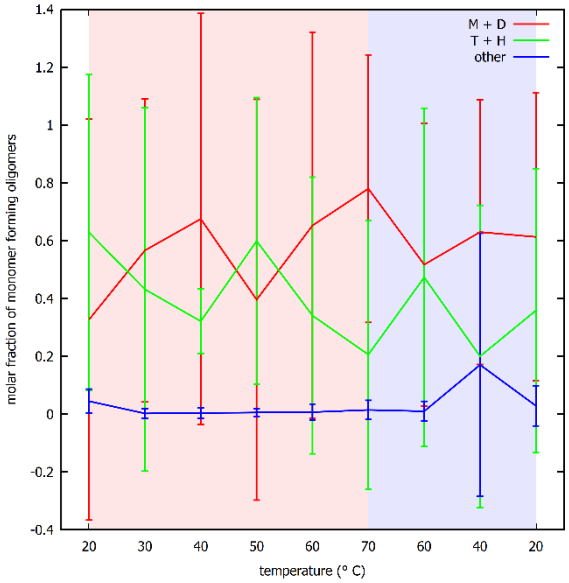
Insulin 2 g/L + EC-212 0.1 M



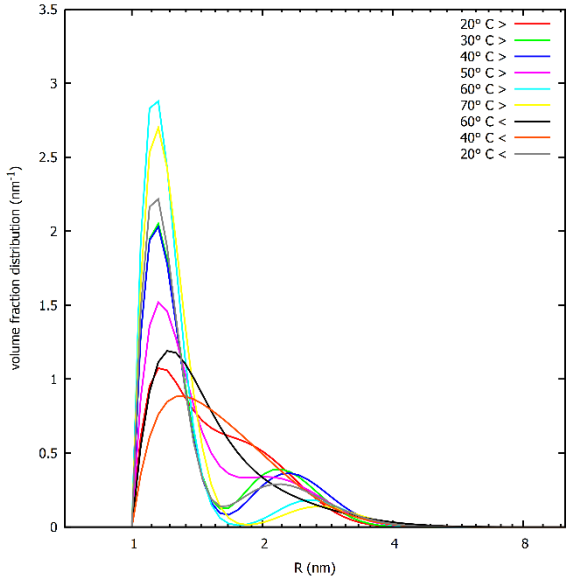


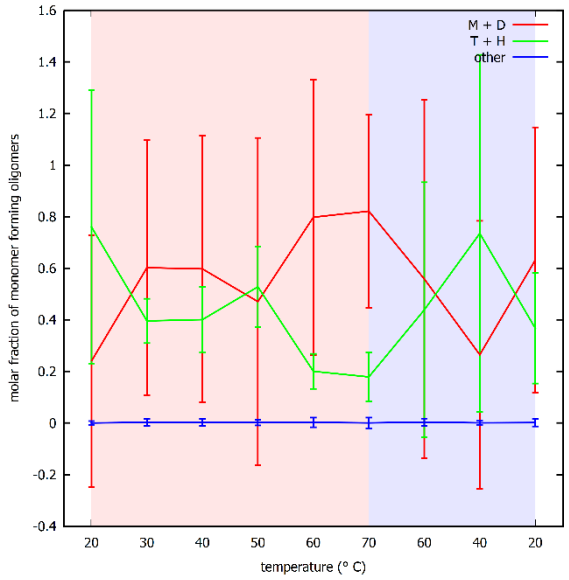
Insulin 2 g/L + EC-212 0.05 M



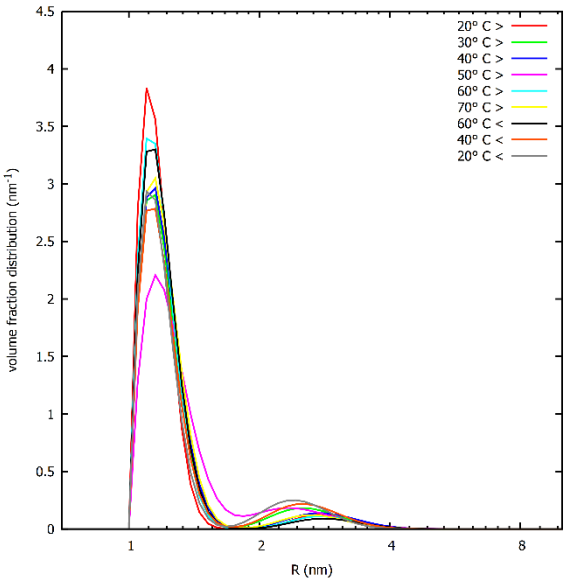


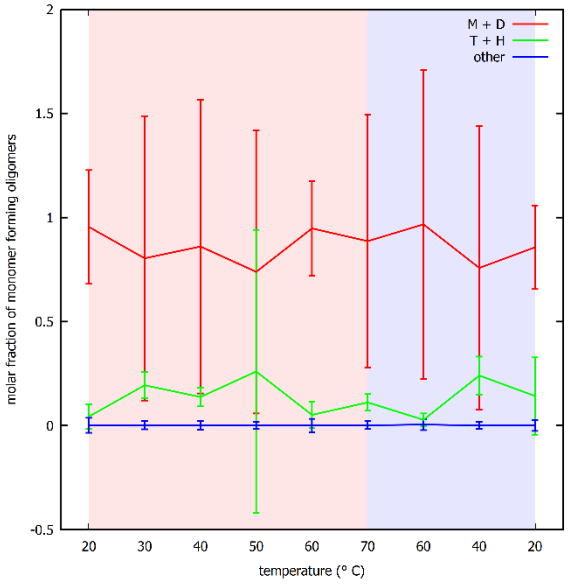
Insulin 2 g/L + EC-312 0.1 M





Insulin 2 g/L + EC-312 0.05 M





References

[1] Heinemann L., Braune K., Carter A., Zayani A., Krämer L. A. Insulin Storage: A Critical Reappraisal. *Journal of Diabetes Science and Technology* 2021, Vol. 15 (I) 147-159.

[2] Mensink M. A., Frijlink H. W., van der Voort Maarschalk K., Hinrichs W. L. J. How sugars protect proteins in solid state and during drying (review): Mechanism of stabilization in relation to stress conditions. *European Journal of Pharmaceutics and Biopharmaceutics 114* (2017), 288-295.

[3] Spinozzi F., Mariani P., Ortore M. G. Proteins in binary solvent. *Biophysical reviews* (2016).

[4] Kamerzell T. J., Esfandiary R., Joshi S. B., Middaugh C. R., Volkin D. B. Protein-excipient interaction: Mechanism and biophysical characterization applied to protein formulation development. *Advanced Drug Delivery Reviews 63* (2011), 1118-1159.

[5] Mari E., Ricci C., Pieraccini S., Spinozzi F., Mariani P., Ortore M.G. Trehalose Effect on The Aggregation of Model Proteins into Amyloid Fibrils. *Life 10* (2020), 60.

[6] Olsson C., Swenson J. The role of disaccharides for protein-protein interactions – a SANS study. *Molecular physics* (2019), 117:22, 3408-3416.

[7] Ajito S., Hirai M., Iwase H., Shimizu N., Igarashi N., Ohta N. Protective action of trehalose and glucose on protein hydration shell clarified by using X-ray and neutron scattering. *Physica B: Condensed Matter 551* (2018), 249-255.

[8] Maycock C. D., Centeno M. R. M. B. V., Lourenço E.C., Dos Santos M. H. D., Miguel A. S. D. C. Hexose Derivatives, Preparation and Uses Thereof. *U. S. Patent Application* (2018) US2018/0170955A1.

[9] 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* (2022), 123.

[10] Wittenberg J. B., Wittenberg B. A. Myoglobin function reassessed. *The Journal of Experimental Biology 206* (2003), 2011-2020.

[11] Kendrew J. C., Bodo G., Dintzis H. M., Parrish R. G., Wyckoff H., Phillips D. C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature 181* (1958), 662-666.

[12] Maurus R., Overall C. M., Bogumil R., Lou Y., Mauk G. A., Smith M., Brayer G. D. A myoglobin variant with a polar substitution in a conserved hydrophobic cluster in the heme binding pocket. *Biochimica et Biophysica Acta 1341* (1997), 1-13.

[13] Kendrew J. C., Dickerson R. E., Strandberg B. E., Hart R. G., Davies D. R., Phillips D. C., Shore V. C. Structure of myoglobin, a three-dimensional Fourier synthesis at 2 Å resolution. *Nature 185* (1960), 422-427.

[14] Frauenfelder H., McMahon B. H., Austin R. H., Chu K., Groves J. T. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Biophysics and Computational Biology 98* (2001), 5, 2370-2374.

[15] Garry D. J., Meeson A., Yan Z., Williams R. S. Life without myoglobin. *Cellular and Molecular Life Sciences 57* (2000), 896-898.

[16] Tokarz V. L., MacDonald P. E., Klip A. The cell biology of systemic insulin function. *Journal of Cell Biology 217* (2018), 7, 2273-2289.

[17] Ward C. W., Lawrence M. C. Landmarks in insulin research. *Frontiers in Endocrinology 2* (2011), 1-11.

[18] Mayer J. P., Zhang F., DiMarchi R. D. Insulin structure and function. *Peptide Science 88* (2007), 687-713.

[19] Coffman F. D., Dunn M. F. Insulin-metal ion interactions: the binding of divalent cations to insulin hexamers and tetramers and the assembly of insulin hexamers. *Biochemistry 27* (1988), 6179-6187.

[20] Braun A. G. Structure of insulin in 4-zinc insulin. *Nature 261* (1976), 166-168.

[21] Goodsell D. Insulin. *PDB-101, https://pdb101.rcsb.org/motm/14*

[22] Lan J., Ge J., Yu J., Shan S., Zhou H., Fan S., Zhang Q., Shi X., Wang Q., Zhang L., Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature 581* (2020), 215-220.

[23] Yao H., Song Y., Chen Y., Wu N., Xu J., Sun C., Zhang J., Weng T., Zhang Z., Wu Z., Cheng L., Shi D., Lu X., Lei J., Crispin M., Shi Y., Li L., Li S. Molecular architecture of the SARS-CoV-2 Virus. *Cell 183* (2020), 3, 730-738.

[24] Yan R., Zhang Y., Li Y., Xia L., Gou Y., Zhou Q. Structural basis of the recognition of SARS-CoV-2 by full-length human ACE2. *Science 367* (2020), 1444-1448.

[25] Salamanna F., Maglio M., Landini M. P., Fini M. Body localization of ACE-2: on the trail of the keyhole of SARS-CoV-2. *Frontiers in Medicine 7* (2020).

[26] Barros E. P., Casalino L., Gaieb Z., Dommer A. C., Wang Y., Fallon L., Raguette L., Belfon K., Simmerling C., Amaro R. E. The flexibility of ACE2 in the context of SARS-CoV-2 infection. *Biophysical Journal 120* (2021), 6, 1072-1084.

[27] Master thesis degree of Doctor Palombini M., relator professor Mariani P. (2021). Recombinant production of ACE2 protein for SARS-CoV-2 detection by sensor.

[28] Malvern Panalytical (2019). Zetasizer Pro and Ultra DLS Customer Training Course.

[29] Bhattacharjee S. (2016). DLS and Zeta potential – What they are and what they are not?. *Journal of Controlled Release*, *235* (2016), 337-351.

[30] European Synchrotron Radiation Facility. What is a synchrotron?. *esrf.fr*. https://www.esrf.fr/cms/live/live/en/sites/www/home/about/synchrotron-science/synchrotron.html (June 2022).

[31] Pernot P., Theveneau P., Giraud T., Nogueira Fernandes R., Nurizzo D., Spruce D., Surr J., McSweeey S., Round A., Felizas F., Foedinger L., Gobbo A., Huet J., Villard C. and Cipriani F. (2010). New beam line dedicated to solution scattering from biological macromolecules at the ESRF. *J. Phys.: Conf. Ser.* 247, 012009.

[32] Round A., Felizas F., Fodinger L., Gobbo A., Huet J., Villard C., Blanchet C. E., Pernot P., McSweeey S., Roessle M., Svergun D. I., Cipriani F. (2014). BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high throughput X-ray solution scattering experiments. *CrossMark, Acta Crystallographica Section D, ISSN 1399-0047.*

[33] Scotti A., Liu W., Hyatt J. S., Herman E. S., Choi H. S., Kim J. W., Lyon L. A., Gasser U., Fernandez-Nieves A. (2015). The CONTIN algorithm and its application on determine the size distribution of microgel suspension. *The journal of chemical physics 142*, 234905.