UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN FISICA

CICLO XXXIII

Investigation of collagen self-assembly mechanisms and dehydration processes by means of MD simulations, UV and FTIR measurements

Coordinatore:
Chiar.mo Prof. Stefano Carretta

Tutore:
Chiar.ma Prof.ssa Eugenia Polverini

Dottoranda:
Dott.ssa Ludovica Leo

Anni Accademici 2017/2018 – 2019/2020

Alla mia famiglia

Index

1. Introduction 2
   1.1 Collagen 2
      1.1.1 Triple helix 2
      1.1.2 Fibrillar collagen 4
         1.1.2.1 Type I collagen 6
   1.2 Collagen fibrillogenesis 6
   1.3 Collagen and water 10
      1.3.1 The importance of hydration shell 10
      1.3.2 Collagen hydration and dehydration 13
   1.4 Aim of the study 15
2. Computational techniques 18
   2.1 Molecular dynamics technique 18
      2.1.1 The force field 20
         2.1.1.1 Bonded interactions 20
         2.1.1.2 Non-bonded interactions 21
         2.1.1.3 The force field choice 22
      2.1.2 Periodic boundary conditions 23
      2.1.3 The solvation 23
      2.1.4 The energy minimization 24
      2.1.5 The MD trajectory 24
      2.1.6 Standard simulation protocol 26
      2.1.7 Limits of MD technique 27
3. Experimental techniques 28
   3.1 UV/visible spectroscopy 28
      3.1.1 Spectral Artifacts: light scattering sample 29
      3.1.1 UV/visible spectrophotometer 30
   3.2 IR spectroscopy 31
      3.2.1 Molecular vibrations 32
      3.2.2 The selection rule 33
      3.2.3 FTIR Spectroscopy 34
         3.2.3.1 FTIR spectrophotometer 35
         3.2.3.2 Application of FTIR spectroscopy to biological systems 36

4. Materials and methods 39
4.1 Molecular dynamics simulations 39
4.1.1 Model building of triple helices 39
4.1.2 Molecular dynamics simulations set-up 41
4.1.2.1 MD simulations set-up details for the study of collagen self-assembly 42
4.1.2.2 MD simulations set-up for collagen dehydration study 42
4.1.3 Methods for the MD trajectories analysis 43
4.1.3.1 Methods for the analysis of MD trajectories: a collagen self-assembly study 44
4.1.3.2 Methods for the analysis of MD trajectories: a collagen dehydration study 44
4.2 Experimental studies 46
4.2.1 UV-visible measurements to study collagen self-assembly 46
4.2.1.1 Materials 46
4.2.1.2 Measurements 46
4.2.1.3 Data analysis 47
4.2.2 FTIR measurements to study collagen dehydration-rehydration cycle 47
4.2.2.1 Materials 47
4.2.2.2 Experimental procedures and measurements 48
4.2.2.3 Spectra analysis 49
5. Results and discussion: study of collagen self-assembly 51
5.1 MD simulations 51
5.1.1 Aggregation of LC fragments 51
5.1.1.1 Physiological salt concentration 51
5.1.1.2 Low ionic strength 57
5.1.2 Aggregation of HC model fragments 59
5.1.2.1 Physiological salt concentration 59
5.1.2.2 Low ionic strength 64
5.1.3 Differences in the aggregation between LC and HC models: insights into the assembly mechanism 66
5.2 UV measurements 69
6. Results and discussion: study of collagen dehydration 72
6.1 MD simulations 72
6.2 Experimental measurements: gravimetric and FTIR studies 81
7. Conclusions 89
References 93

1. Introduction

1.1 Collagen

1.1.1 Triple helix

Collagen is a rigid fibrous protein, component of extracellular matrix in animals and main constituent of bones, cartilage, skin and tendon [1]. A right-handed triple helix structure is the basic structural unit of collagen and it is named tropocollagen. Tropocollagen is composed by intertwined left-handed helices, about thousand residues long, forming the characteristic poly-Pro II-like triple helix [1].

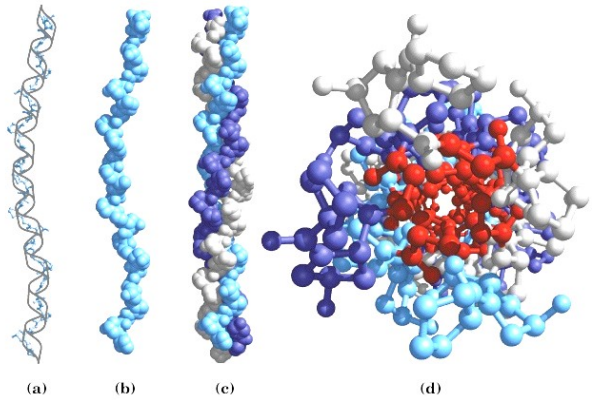


Figure 1. Tropocollagen molecule: single chain representation (a), (b), triple helix along its axis (c) and triple helix in section view (d). Adapted from [2].

The primary structure of tropocollagen is very peculiar due to the characteristic repetition of Gly-Xaa-Yaa triplets, where one-thirds of Xaa and Yaa are proline or 4-hydroxyproline (Hyp) residues. The latter amino acid derives from a post translational modification of proline residue, in which a hydroxyl group is linked to the fourth carbon. This characteristic collagen sequence, rich in glycine and in imino acids with a rigid cyclical structure (Hyp and Pro), combined with electrostatic interactions between aspartate and lysine, assures stabilization of the tropocollagen [3]. Hydroxyproline is not the only modified residue in collagen molecule, because also a post-translational modification of Lys occurs, giving rise to 5-hydroxylysine.

These two modified residues are formed after collagen synthesis, through the action of prolil-hydroxylase and lysil-hydroxylase enzymes [4].

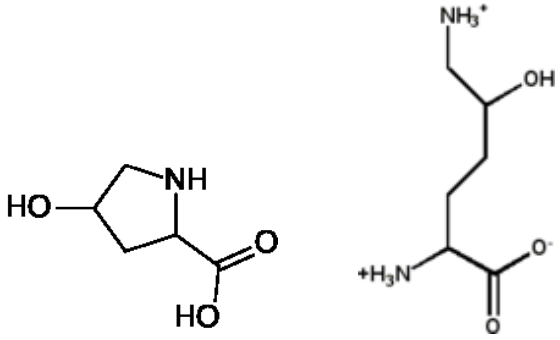


Figure 2. 4-hydroxyproline (on the left) and 5-hydroxylysine (on the right)

Several studies have demonstrated that hydroxyproline stabilizes the triple helix through a stereoelectronic effect of the electronegative oxygen, arranging the main chain in the correct conformation for triple-helix formation. Instead, Hyp hydration contributes little (if any) to triple helix stability [5]. Some hydroxylysines are further modified by O-linked glycosylation, that is, galactose and glucose are attached to the hydroxyl group of hydroxylysine. Alterations in the levels of collagen glycosylation have been reported in several bone/skeletal disorders, suggesting a role of the glycosylation in bone mineralization [6].

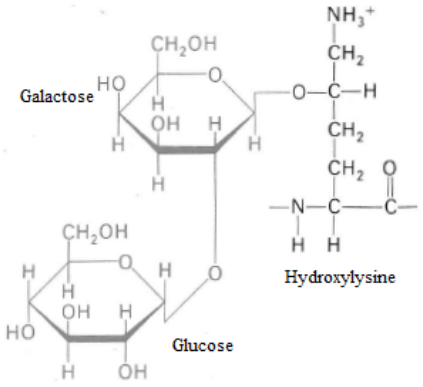


Figure 3. Collagen glycosylation. Adapted from [7]

The high presence of Gly, Pro and Hyp makes impossible the formation of other secondary structures like alpha-helix or beta-sheet. Triple helix is more narrow and extended compared to alpha-helix, with an extension for residue along helix axis of 0.29 nm against 0.15 nm of alpha-helix [4].

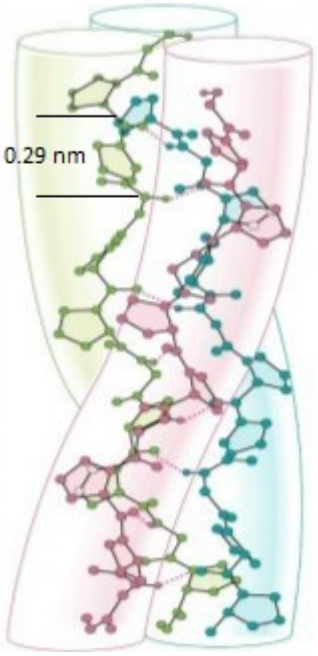


Figure 4. Poly-Pro triple helix, adapted from [4].

In a triple helix, one residue every three overlooks and contacts the inside of structure, which is the point of closest approach among the three chains. Only a small residue, like glycine, can stay in this crowded zone, so, for this reason, the presence of Gly-Xaa-Yaa triplets is fundamental for the triple helix formation. The three chains are staggered so that glycines of a chain are close to Xaa of other chain and to Yaa of the third one. This arrangement enables amino groups of glycines to make hydrogen bonds with carboxylic groups of the other residues [4].

1.1.2 Fibrillar collagen

About 30 distinct collagen types have been identified in vertebrates, so far [1]. They can be divided into two categories: fibrillar and non-fibrillar collagen. The former is able to form fibril structures and includes type I collagen, the most common one, prevailing in bones, tendons and skin. Other frequent fibrillar collagens are the type II, present in cartilage and the type III, prevalent in blood vessels.

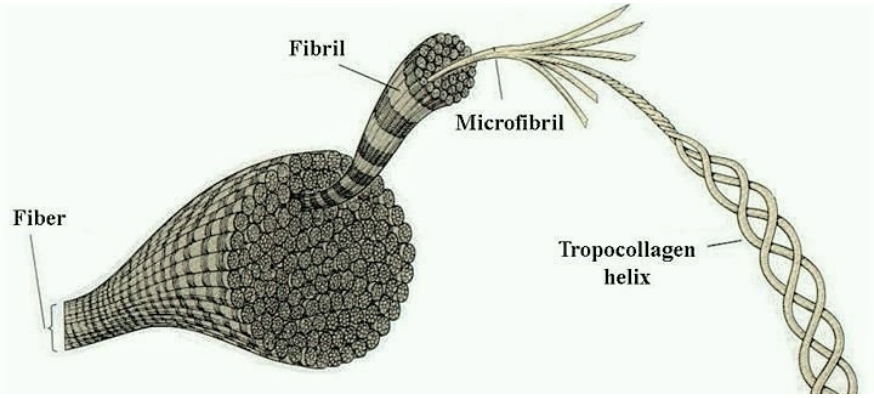


Figure 5. Schematic structure of fibrillar collagen. Adapted from [8]

Fibrillar collagens can be homotrimers or heterotrimers [9]. The genes encoding the chains of a triple helix (named alpha chains) derive from a common ancestor formed by 54 or 45 nucleotides, encoding 18 or 15 residues, respectively. This ancestor, through duplication and mutation events, gave rise to larger genes that codify for a transcript of a collagen monomer in a range of 1200–1300 amino acids [10]. It means that each chain of fibril collagen composing a tropocollagen, is formed by an uninterrupted sequence of about 300 (Gly–Xaa–Yaa) triplets, with terminal domains loosing triple helix structures, the so-called telopeptides. These fibrillar collagens are rich in lysine, glutamine and arginine, as in the ancestor sequence [10].

The mature structure of tropocollagen derives from its precursor, the procollagen, through the action of proteinase. Fibrillar procollagens include the triple helix domain, the telopeptides domains and two terminal sites, named N-propeptide and C-propeptide. These C- and N- propetides are not triple helix structures, they are composed by 30 and 33-54 residues, respectively, and prevent a premature fibrillogenesis. Only after the proteinases have cut these terminal sites, the procollagen becomes tropocollagen and fibril formation can occur. Fibrillar tropocollagens have a molecular weight of about 285 kDa, and have a length of about 300 nm and a diameter of 1.4 nm [1].

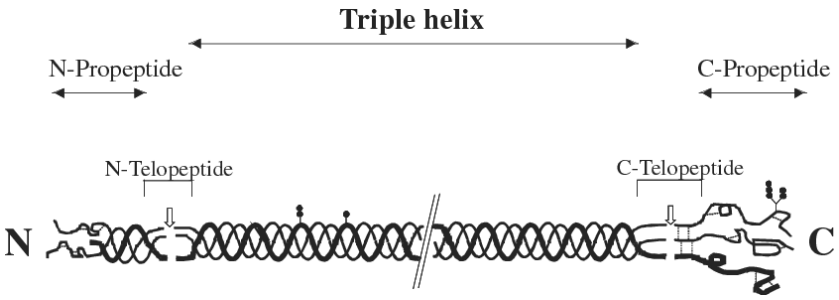


Figure 6. Schematic structure of a procollagen molecule, adapted from [11].

1.1.2.1 Type I collagen

Type I collagen is the most abundant in tissues and is the protein used in this research work. The triple helix consists of two identical polypeptide chains, named α1, and of a different one, named α2. These two types of α chains derive from pro-α1 and pro-α2 precursors, which have a molecular weight of 120 kDa. The additional C- and N-propeptides present in pro-α1 and pro-α2 have a completely different aminoacidic composition as compared to the triple helix portion of α1 and α2 chains. In fact, they are not rich in glycines, prolines and hydroxyprolines, and they form cross-links between chains through disulfide bridges that are absent in α1 and α2 chains [7].

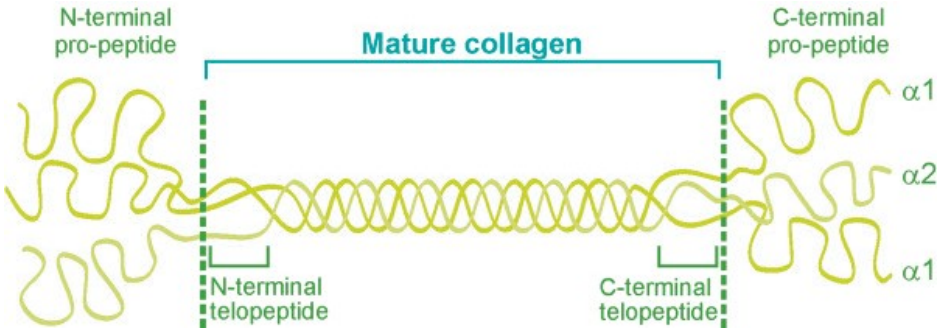


Figure 7. Schematic structure of Type I pro-collagen. Adapted from [12]

1.2 Collagen fibrillogenesis

After the cleavage of propeptides, tropocollagen molecules are ready to assembly and form fibrillar structures. The fibrillogenesis process is led by loss of water from the

surface of monomers and is defined as an entropy-driven process [10]. The supramolecular structures that are formed during fibrillogenesis are initially preserved by non-covalent interactions (e.g. hydrophobic or polar contacts). Afterwards, cross-links between lysine sidechains of telopeptides and triple helix region are formed, to increase the stability of aggregates [10] [13]. In particular, lysine sidechains undergo aldolic condensation, forming cross-links among tropocollagens. The flexibility of telopeptide structures facilitates the formation of these cross-links [7]. Quantity and type of cross-links change depending on physiologic function and tissue age. For example, the collagen of Achille's tendon in an adult rat has got a high number of cross-links, in contrast to collagen of flexible rat tail tendon, that has got a lower number [7]. Since covalent interactions are established only on mature fibrils, telopeptides could only speed up the assembly process, but they are not crucial to initiate the fibrillogenesis [14].

The primary structure of tropocollagen molecule shows charged and hydrophobic residues allocated with a certain periodicity. Due to this periodicity, a characteristic stacking of neighboring tropocollagens is required in the fibril [10]. Transmission electron microscopy experiments showed a characteristic banded structure in collagen fibrils, resulting in a regular array of overlap and gap regions caused by a periodic staggering (named D-spacing) between neighboring collagen monomers [14].

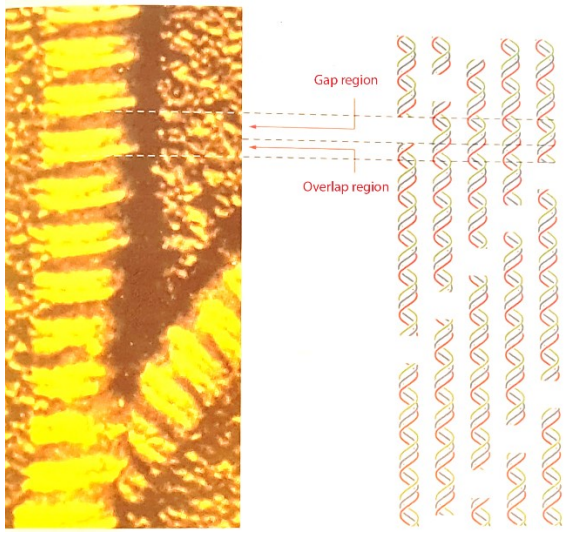


Figure 8. Banded structure of collagen recorded by TEM (on the left) composed by gaps and overlaps of tropocollagens molecules (schematically represented on the right). Adapted from [4].

D-spacing assumes different values depending on type of collagen and tissue. A distribution of D-spacing values has been reported by Fang and Holl [15](Figure 9).

Table 1 D-spacing variations reported in literature

Variations in D-spacing	*Tissue and technique*	*Results and discussion*
55–80 nm distribution 64.6 ± 5.3 nm[a]	Human skin TEM	The 55–80 nm range of spacing is not unique to collagen, but also shared by neurotubules
64.6 ± 0.8 nm in cornea 67.7 ± 0.8 nm in tendon	Cornea and tendon XRD	An 18° axial inclination in the cornea explains the D-spacing difference between the cornea and the tendon ($\cos\alpha = Dc/Dt$)
67.7 ± 0.9 nm in central zone 71.3 ± 0.4 nm in distal zone	Vitrified predentin TEM	D-spacing differences in the two zones may be due to the presence of proteoglycans and ions that bind to collagen
54–75 nm distribution (predominantly 67–68 nm hydrated; 57, 62, 67 nm dehydrated)	Partially demineralized dentin AFM	Reduced D-spacing may be due to dehydration-induced structure disorder and loss of crystallinity
69.6 ± 2.9 nm	Rat digital tendon AFM	Fibril D-spacing is preserved independent of the fibril diameter
63–73 nm distribution	Mice bone, dentin and tendon AFM	A distribution of D-spacing exist in the bone, dentin and tendon, regardless of the presence of mineral, cellular origin, anatomical location or mechanical function of the tissue
68.0 ± 2.6 nm in sham; 65.9 ± 3.1 nm in OVX	Sham and OVX ovine radius bone AFM	Estrogen depletion induces changes in type I collagen nanomorphology of bone ($P < 0.001$)
63–74 nm distribution in WT; 56–75 nm distribution in brtl/+	WT and brtl/+ mice femur bone AFM	D-spacing means between WT and brtl/+ are not different (67.6 nm vs 67.4 nm); D-spacing distributions between the phenotypes are statistically different ($P = 0.001$)
59–66 nm distribution in sham; 56–67 nm distribution in OVX	Sham and OVX ovine dermis AFM	Estrogen depletion induces changes in type I collagen nanomorphology of dermis ($P < 0.001$)
57–69 nm distribution	Ovine bone AFM	Fibrils from one D-bundle share similar D-spacing; a distribution of values arises at the bundle level

Abbreviations: AFM, atomic force microscopy; OVX, ovariectomized TEM, transmission electron microscopy; WT, wild type; XRD, X-ray diffraction.

Figure 9. D-spacing distribution reported by Fang and Holl. Adapted from [15].

Alterations in bone collagen D-spacing values are observed in some diseases, like Osteogenesis Imperfecta, or disorders related to long-term estrogen depletion [15]. The most reliable value of D-spacing reported in literature for collagen type I in tendons and bones, is 67 nm. It has been demonstrated that maximal hydrophobic and electrostatic interactions take place when tropocollagens are shifted by 234±1 amino acids along the fibril axis, which roughly corresponds to 67 nm [15]. In particular, type I collagen sequence can be divided into five segments, with dissimilar charge profiles, where four segments have the same length, namely D. The fifth segment is 0.4 D long. When type I collagen assembly occurs, adjacent tropocollagens are shifted by 1 D, whereas longitudinal neighboring molecules are separated by gaps of 0.6 D. Therefore, the characteristic banded structure observed by TEM in type I collagen derives from overlaps of 0.4 D and gaps of 0.6 D [10]. It

has been observed that *in vitro* assembly of type I collagen provides similar D-spacing values, suggesting that the typical staggering is intrinsic to the fibrillar collagen self-assembly, that, hence, does not require interfibrillar crosslinking or sugar binding [15].

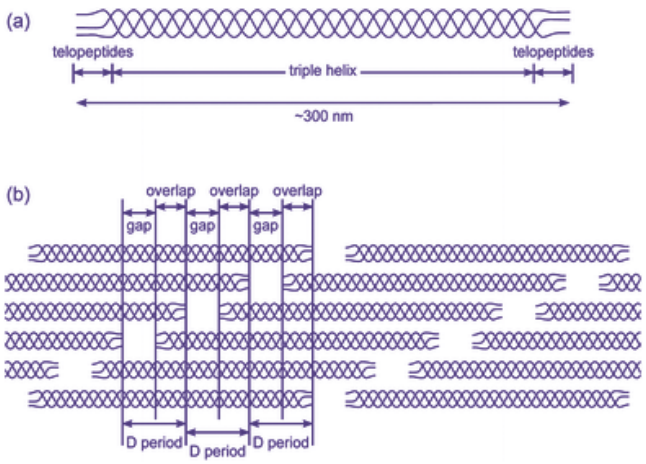


Figure 10. Schematic representation of tropocollagen (a) and fibril (b) structures. D period, gap and overlap regions are highlighted. Adapted from [16]

The thinnest collagen aggregate reproducing this characteristic structural pattern of gaps and overlaps is microfibril. Microfibril is composed by five staggered tropocollagens arranged in a cylindrical shape that is 4-5 nm in diameter [10].

A collagen fibril is composed by microfibrils and can reach 1 µm in length. The lateral association of fibrils gives a fiber, the largest supramolecular structure of collagen. Collagen fiber can reach 10 µm in length and from 10 to 300 nm in diameter [17].

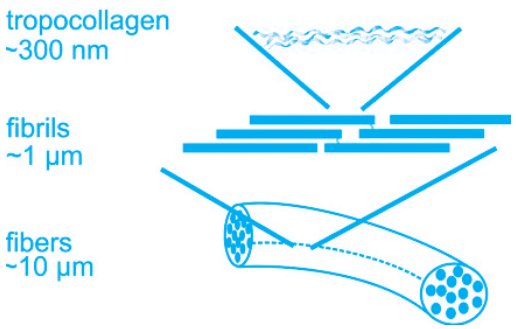


Figure 11. Schematic structures of tropocollagen, fibril and fiber. Adapted from [17].

In vivo fibrillogenesis takes place in extracellular matrix, as shown in the following picture (Figure 12), displaying the main steps of collagen synthesis.

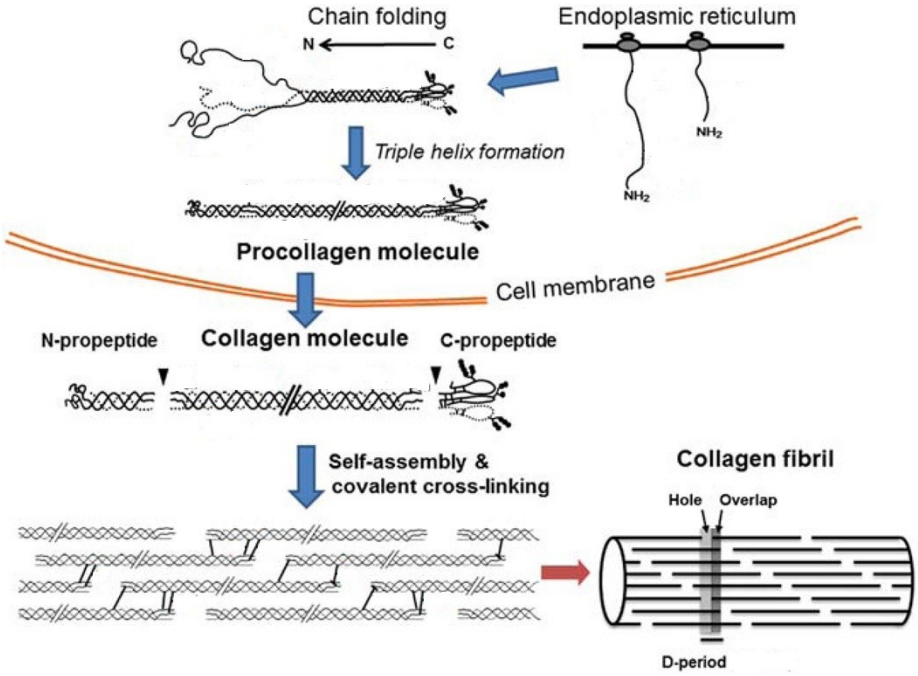


Figure 12. Schematic representation of collagen biosynthesis steps. Collagen fibrils are formed in extracellular matrix, after the cleavage of propeptides. Adapted from [6].

1.3 Collagen and water

1.3.1 The importance of hydration shell

Water is a fundamental medium for biological macromolecules. Water molecule is a permanent dipole able to participate up to four hydrogen bonds. In liquid water, the tridimensional network of water molecules is not uniform due to local rearrangements in H-bonds. The average lifetime of an hydrogen bond between two water molecules is 9.5 ps, and, therefore, the time in which a water molecule moves, re-orients and interacts with other neighboring water molecules is about 10 ps [4].

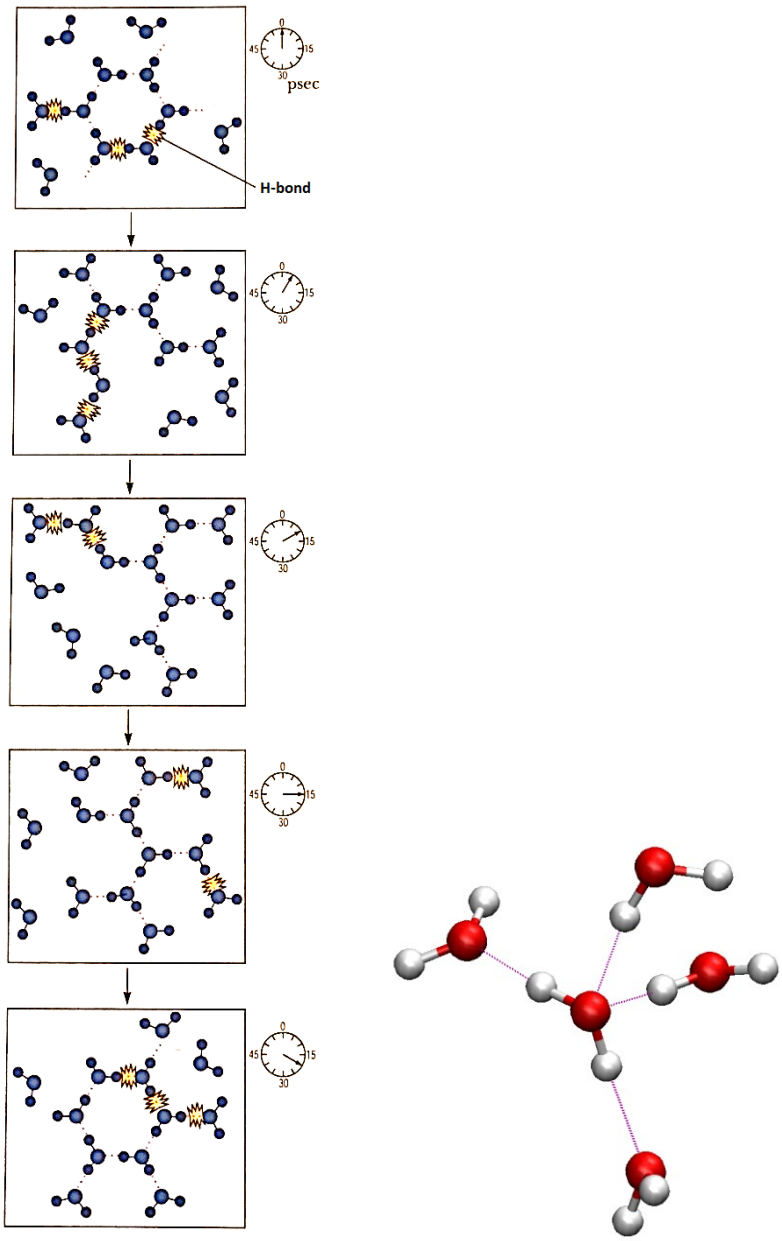


Figure 13. **(On the left) Network of water molecules in liquid phase. Formation and breaking of H-bonds are highlighted. Adapted from** [4]**. (On the right) H-bonds that a water molecule is able to form with surrounding water molecules.**

Water is an excellent solvent for polar substances, due to its dual function of donor and acceptor of hydrogen bonds. The closest water around biological macromolecules is called hydration water or hydration shell. In a simplified model of a protein, water molecules surrounding it create ordered structures around the polypeptide chain with different binding energies and different mobility. The outer shell is formed by water molecules whose mobility is similar to that of liquid water and it is blended in the bulk water.

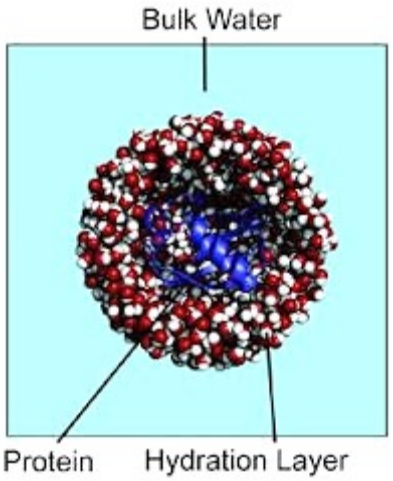


Figure 14. Schematic representation of hydration shell around a generic protein. Adapted from [18]**.**

In protein systems inner water is very important to preserve native structure and function: for example, in globular protein, some water trapped inside the cavities and the deep grooves of the polypeptide chain are necessary to stabilize the protein structure or to play a particular role.

The functional role of hydration water in protein systems was discovered early [19]. Hydrodynamic measurements were used not only to define the shape and size of proteins but also to perform analyses focused on evaluation of the solvent amount around them. Studies on partially hydrated powders has allowed to examine the interactions between hydration water and protein surface [19]. Hydration can be defined as a process in which water is incrementally added to dry protein until no more changes are induced to protein structure by addition of other water molecules. This hydration level can be defined as hydration end point [19]. In light of these considerations, the hydration shell can be described as the amount of water associated with the hydration end point of protein. It was demonstrated that the hydration shell corresponds to a monolayer that covers the whole protein surface. To identify the level of hydration, the mass ratio (grams of water per grams of dry protein) is commonly adopted and indicated as h unit [19]. However, it has to be considered that hydration of protein depends on solvent-accessible surface area rather than weight. h values related to the hydration end point can be different depending on the type of protein. For example, studies on lysozyme powders have identified 0.38 h as the hydration level in which the water monolayer that covers lysozyme surface is locally ordered around polar and charged groups [19].

Studies on hydration processes of enzymes were carried out to investigate how hydration can influence the enzymatic activity [19]. Enzyme powders, like lysozyme, exhibit a critical hydration level for the onset of enzymatic activity, that is absent at very low hydration levels. For example, chymotrypsin and other enzymes develop their activity starting from 0.1-0.15 h, while below this threshold enzymatic function is negligible. This critical hydration level is considered a crucial condition for the flexibility of the protein-water system that is fundamental for the protein function [19]. In fact, at very low hydration levels, water mobility is low and protein motions are essentially frozen. In this condition, the large-scale collective motions are hindered and the only allowed dynamics is the one relative to picoseconds or sub-picoseconds time scales, leading the protein in a sort of glassy state. Upon hydrating, the water acts as a catalyst, promoting conformational transitions by participating in H-bonds network, so reducing the free energy of the system.

It has been reported that a possible unfolding induced by low hydration is not revealed at accessible temperature [19]. Nevertheless, dehydration at low hydration level can induce structural changes in proteins. It has been demonstrated that lyophilization (freeze+drying) may induce changes in IR spectra of protein [20]. In particular, conversion of helix to beta-sheet during lyophilization was observed from deconvolution of Amide I in IR spectra of many proteins. The increase in ß-sheet content is often associated with protein aggregation, because of the increasing of interprotein interactions. Considering that lyophilization causes the removal of a part of the first hydration shell, the increase in the number of inter-protein H bonds is due to the interactions between facing polar groups, replacing the previous ones with water. For this reason, proteins tend to aggregate during lyophilization process [20].

1.3.2 Collagen hydration and dehydration

Water plays a substantial role in the stability of collagen structures, because it improves helices flexibility by facilitating fibrillar alignment and recognition, through mediated hydrogen bonds [3]. It provides water bridges participating in inter-chain hydrogen bonds and forming a cylinder of hydration around collagen surface, as observed by Bella and coworkers from studies on a collagen peptide [21].

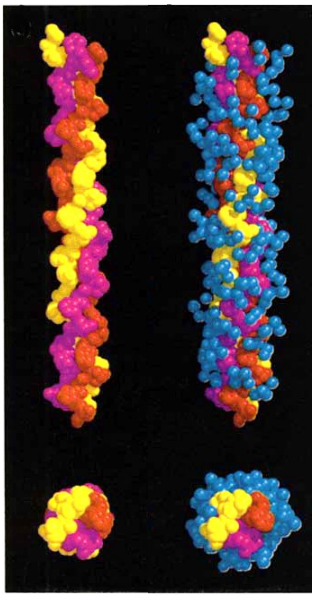


Figure 15.Collagen peptide without (on the left) and with the cylinder of hydration (on the right). Adapted from [21].

Water lacking induces irreversible damages in collagen conformation. Water depletion can occur in various events, like some diseases as diabetes, UV light exposition and aging process. These pathologies originate oxidation products, so-called Advanced Glycation End-products (AGE), due to glycation reactions of collagen molecules with carbohydrate matrix [3]. The "Advanced" word indicates that its formation takes place following a series of reactive intermediates, going through rearrangement, oxidation, dehydration, and fragmentation reactions of sugar-protein adducts [22]. The AGE products accumulate with aging [23], modifying physical properties of fibers, such as increase in fiber diameter and stiffness, so causing a loss of flexibility and solubility of tissues [3]. Moreover, it has been observed that dehydration itself of collagen fibers below certain threshold values induces an irreversible loss of their flexibility. In this condition, collagen become a rigid material with different proprieties as compared with its native state [24]. The loss of flexibility is not the only change induced by dehydration in collagen molecules, but also modifications in fibrils arrangement can occur. In fact, an IR study on collagen film [3] has pointed out an increase in ß-sheet content, suggesting the rising of new interactions between neighboring fibrils inducing a packing, as in lyophilization process.

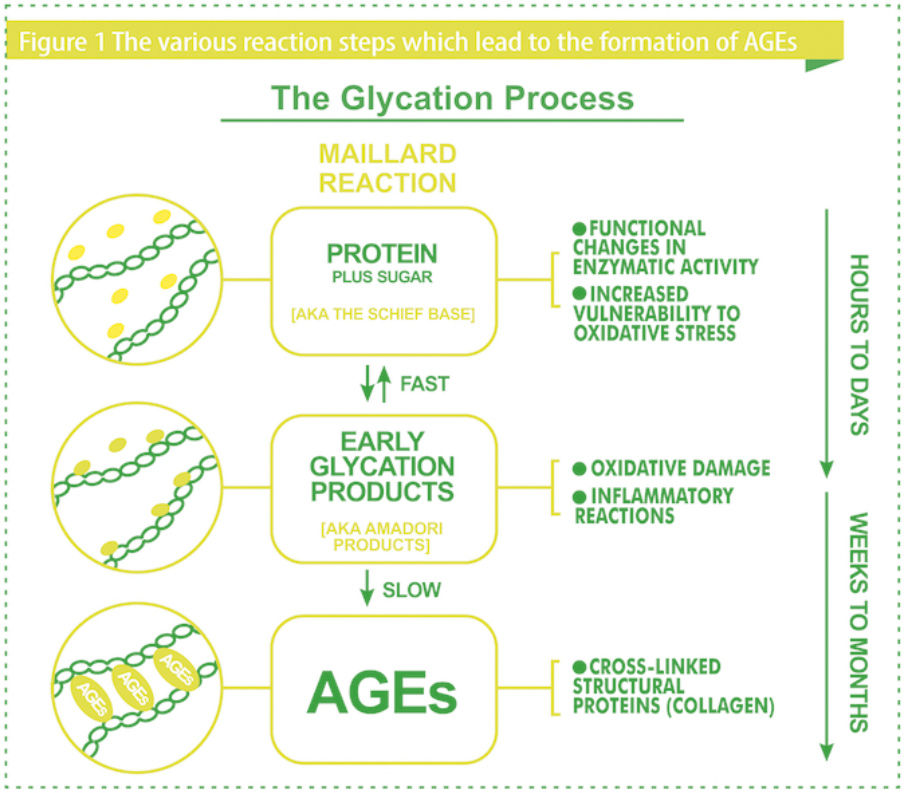


Figure 16. The glycation processes. Adapted from: www.prime-journal.com

1.4 Aim of the study

In this work two main studies on collagen molecules have been carried out: collagen self-assembly mechanisms and collagen dehydration processes. The use of different techniques to investigate these phenomena enables to capture different details and to infer information on the behavior of the protein in vivo.

We have studied the collagen self-assembly process by coupling UV spectroscopy and molecular dynamics simulations in order to explore various aspects of the problem. The former technique enabled to monitor the process for the whole fibers in different physicochemical conditions, whereas the latter one allowed to focus at atomic level on the aggregation of four small fragments, observing molecular interactions and behavior. Moreover, variations in the self-assembly mechanism was detected changing the solution proprieties. We have monitored the fibers formation at three different pH values (5.0, 7.0 and 8.0), close to physiological pH, by means of UV absorption technique in order to discover the optimal pH for collagen aggregation. It

has been demonstrated that collagen self-assembly does not occur at extreme pH values because of the repulsive forces of the highest net charge [25], so, for these reasons, we have limited the study only at pHs slightly below or above neutrality. We followed the methods of Nomura and co-workers [26], monitoring the aggregation process through the turbidity changes of the solutions. Some parameters measuring the kinetics of aggregation, calculated from turbidity curves, enabled to compare the assembly processes at different pHs.

Molecular dynamics simulations have been particularly useful to examine the influence of hydropathic features of the different regions of the chain in collagen self-assembly. This technique seemed to be a very good method to evaluate the interactions established between neighboring tropocollagens during the assembly and to observe the aggregation at atomic level, permitting to infer the behavior of the full-length chain. The literature reported several MD simulations studies on inter-tropocollagen interactions in an already assembled fibril [16] [27] or on the aggregation of especially designed collagen-like peptides [28] [29]. In our case, the actual sequence of type I rat tail collagen was used to build four fragments, staggered of the typical D-spacing of 234 residues, and to monitor their aggregation. The tropocollagens were initially placed at a distance of 2 nm from one another in a box of solvent. We performed simulations not only at physiological salt concentration (0.1 M), but also at low ionic strength in order to investigate the driving force of collagen aggregation mechanism.

The study of dehydration effects on collagen molecules was performed by coupling molecular dynamics and FTIR spectroscopy. Bearing in mind from literature that the water depletion causes damages in collagen molecules, it is interesting to investigate which structural changes are induced by certain levels of dehydration. In addition, it is worth monitoring whether they are reversible or irreversible after a rehydration process and how it is dependent on hydration value reached in the dehydration process. To this aim, we have performed MD simulations on the previously aggregated collagen microfibrils, subjected to dehydration conditions up to the first layer of water and then fully rehydrated. Several MD simulations studies on various protein systems in different hydration conditions (some of them including also other kind of biomolecules) are described by literature [30] [31] [32] [33] [34], especially focused on the stability of a collagen-mimic peptide in vacuo as compared with the hydrated state [24], or on nanomechanics of collagen microfibrils in vacuum

or solvated [35] [36] [37]. The main goal of this study is, rather, to examine the influences of a dehydration up to the first hydration shell on collagen microfibrils structure, with particular regard to their different hydropathic profiles. The comparison of the results obtained from simulations in hydrated, dehydrated and rehydrated state enables to investigate the structural changes induced in the assemblies in dehydration conditions and their reversibility after rehydration. The water mobility and distribution is also monitored in the three conditions.

FTIR spectroscopy combined with gravimetric measurements enables to follow a whole cycle of dehydration-rehydration on whole collagen molecules. To this aim, a gravimetric study on collagen sample has been carried out to investigate which is the water amount desorbed or adsorbed by the sample by varying the hydration conditions. We have performed a parallel FTIR study on collagen sample to monitor the structural changes induced by a dehydration-rehydration treatment through Amide I band deconvolution. Moreover, from FTIR spectra water behavior was also recorded, by water sorption isotherms measured on OH stretching band, and the result was compared with that obtained from the gravimetric study, according to the work of Bridelli et al [3].

2. Computational techniques

Beside the experimental methods, computational approaches have taken an increasingly large part in the studies of biomolecules. These types of techniques is based on energetic calculations and forces evaluation at atomic level. Among them, it is worth citing the molecular dynamics technique. The aim of a molecular dynamics simulation is to provide details on conformational and structural changes of macromolecules subject to specific macroscopic conditions. Most of these simulations are applied to protein studies in order to understand and provide detailed information on their biological functions. Several MD software packages have been implemented, e.g. GROMACS [38], AMBER [39], CHARMM [40] and NAMD [41], and the first one has been used to perform the MD simulations described in this thesis.

The advantages of the molecular dynamics technique with respect to experimental methods regard:

- spatial resolution → the Å scales are reached;
- temporal resolution → the very first events (picosecond scales) are monitored;
- energetic resolution → atomic interaction energies are determined.

In this chapter we provide a detailed description of the molecular dynamics technique.

2.1 Molecular dynamics technique

A molecular dynamics simulation provides detailed information at microscopic level, generating a trajectory of atomic motions of a macromolecule (or a group of macromolecules) in a defined macroscopic state. This technique is based on statistical mechanics concepts by which macroscopic quantities (like temperature, pressure…) are related to microscopic ones (like atomic coordinates, momentum…). In statistical mechanics a microstate is defined as a point in the 6N dimensional *phase space*, in which position **q** and momentum **p** vectors are assigned to each of the N particles belonging to the whole system. All microstates belonging to the same

system, satisfying the same macroscopic conditions, form a set called *ensemble*. In an experiment, a macroscopic sample is an *ensemble*, because it contains a large number of configurations of the system with definite macroscopic proprieties. A molecular dynamics simulation is a temporal evolution of a system whereby it explores the phase space, pass through the microstates corresponding to the different configurations of the system itself. In order to determine the physical quantities, *ensemble* averages are used in both cases: in an experiment the average on all replicas of the system is calculated, while in a MD simulation a time average is computed. In this regard, it is worth citing the ergodic hypothesis, stating that, for a long period of time, the two means coincide:

$$< A >_{ensemble} = < A >_{time}$$

Depending on macroscopic proprieties, different *ensembles* exist and some of them, corresponding to particular physical requirements, were defined:

- microcanonical *ensemble* (NVE), in which the number of atoms N, volume V and energy E are fixed.
- isobaric-isotherm *ensemble* (NPT), in which the number of atoms N, pressure P and temperature T are fixed;
- canonical *ensemble* (NVT), in which the number of atoms N, volume V and temperature T are fixed.

The last two *ensembles* have been used in this work.

The classical molecular dynamics is based on classical potentials, which describe the interactions between the particles of a system, and on the solution of Newton's equations of motion. These systems are composed by atom-like particles, in which nucleus and electrons are kept together and to whom partial charges are assigned knowing the bond polarity. Bonds are treated as springs connecting atoms.
In order to obtain a trajectory for each particle, we need to solve equations of motion, knowing the potential energy of the system:

$$-\frac{dV}{d\vec{r_i}} = m_i\vec{a_i} = m\frac{d^2\vec{r_i}}{dt^2} \qquad (1)$$

where $\vec{r_i}$ and $\vec{a_i}$ vectors are the atom positions and accelerations, and V is the potential energy.

In the following paragraphs we are going to describe the main features of molecular dynamics technique.

2.1.1 The force field

The more common expression of the potential energy of an atomic system is:

$$V = \sum_{bonds} K_b\,(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{torsions} A(1 + \cos(n\tau - \varphi)) + \sum_{torsions} K_\xi(\xi - \xi_0)^2 + \sum_{i,j} -\frac{A}{r_{ij}^6} + \frac{B}{r_{ij}^{12}} + \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \quad (2)$$

The potential energy expression and its parameters constitute a force field. The parameters depend on the type of atoms and on their bonding configurations, and are derived from experiments, quantum mechanics calculations, or both. The potential energy function is the sum of two contributes: the bonded interactions and the non-bonded interactions.

In the next two sub-paragraphs, bonded and non-bonded interactions are described in detail.

2.1.1.1 Bonded interactions

The intramolecular bonded interactions regard atoms connected by one bond (bond stretching), two bonds (angle bending) and three bonds (dihedral angle rotation) and include the first four terms in the equation (2).

The first and second terms are both described by a harmonic potential and regard the stretching and the bending energy, respectively: $E_{stretching} = \sum_{bonds} K_b\,(r - r_0)^2$ and $E_{bending} = \sum_{angles} K_\theta(\theta - \theta_0)^2$. K_b and K_θ are the elastic constants and are related to the stiffness of the bond-spring or the angle-spring, r_0 is the equilibrium distance, and θ_0 is the equilibrium angle. The harmonic potential model for stretching energy differs from reality when r draws near the dissociation point, because in the model bonds cannot break.

The third and fourth terms are related to the torsions around the bond, namely dihedral and improper angles, respectively. The torsional energy concerning dihedral angles, $E_{dihedral} = \sum_{torsions} A(1 + \cos(n\tau - \varphi))$ is described by a periodic function, where τ is the torsion angle, n is the periodicity, A is the amplitude of sinusoidal curve and φ is the phase angle. The torsional energy regarding improper

angles, i. e. torsion angles where the four atoms are not bonded sequentially, $E_{improper} = \sum_{torsions} K_\xi(\xi - \xi_0)^2$, is described again by a harmonic potential that models the out-of-plane bending.

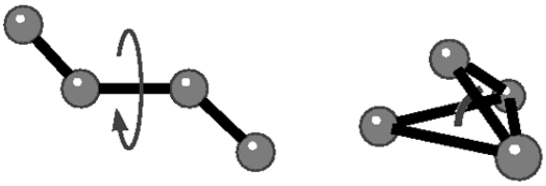


Figure 17 Dihedral angle (left panel) and improper angle (right panel)

2.1.1.2 Non-bonded interactions

The non-bonded interactions between two atoms comprehend Van der Waals and electrostatics interactions.

The first ones are defined by using the Lennard-jones potential: $E_{VdW} = \sum_{i,j} -\frac{A}{r_{ij}^6} + \frac{B}{r_{ij}^{12}}$. The first term describes the attractive contribute, achieved at medium long distances (about > 4 Å), the second one characterizes the repulsive forces, established at short distances (about < 4 Å).

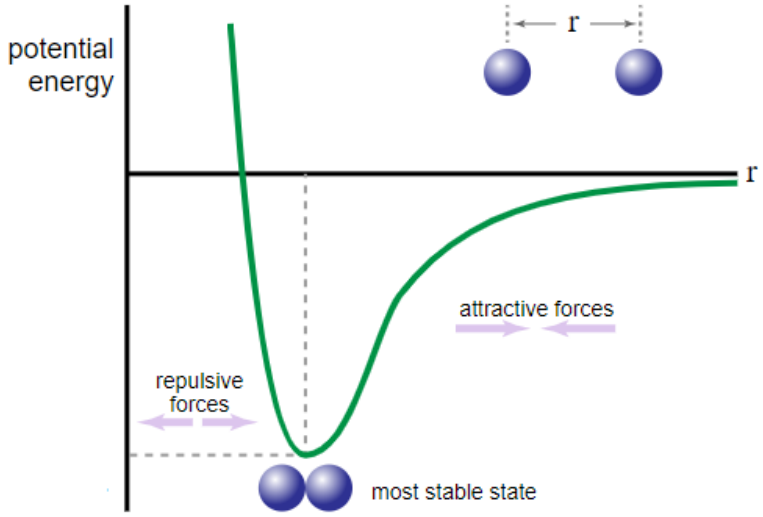


Figure 18 Lennard-Jones potential. Adapted from: https://ch301.cm.utexas.edu/section2.php?target=imfs/mo/mo-geometry.html

The electrostatic interactions are defined by Coulomb potential: $E_{eletrostatic} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}$.

The hydrogen bonds are sometimes implicitly included in the first one of the two non-bonded contributes, while in some force fields they are described by an explicit term.

Non-bonded energy cut off

Several methods of cut off are used with the aim to make the calculation of these energetic contributes faster.

Regarding the Lennard-Jones potential calculation, three methods are mainly used: *truncation*, *shift* and *switch*. The *truncation* method consists in fixing a cut off distance above which the potential drops down to zero. In the *shift* method the potential function is shifted in such a way to become zero at the fixed cut off distance. In the *switch* method the potential expression is modified by a so-called switched function exclusively in a short interval of distances in which the function tends to zero.

Regarding the electrostatic potential calculation, methods based on Ewald summation are the most common. The summation of electrostatic terms is moved from real space to Fourier space, in which it converges more quickly, preventing the loss of fundamental terms when the truncation is operated. However, when the system is big, this method is a too time-consuming process, because the computational cost of Ewald summation increases as N^2. In order to reduce the time computing, an improvement of the Ewald summation was introduced, namely *Particle-Mesh Ewald* (PME) method [42], which uses the Fast Fourier Transform (FFT) to the calculation of the long-range contributions.

2.1.1.3 The force field choice

The force field choice is of crucial importance to a reliable simulation. Several force fields have been developed over the years, with a similar energy expression but different parameters attribution. Among the most used force field families, it is worth to mention GROMOS [43], AMBER [44], CHARMM [45] and OPLS-AA [46], optimized principally for proteins and with a few parameters for organic molecules. GROMOS, which is used in our simulations, is a so called united-atom force field, that is it does not treat non-polar hydrogens explicitly, but incorporates them with carbon atoms assigning the appropriate parameters. The other above-cited force fields are all-atom ones, meaning they treat all hydrogen atoms explicitly. The choice of the force field depends on the systems investigated, so a force filed can be better

than another one for a specific system but can be worse for other systems. This concerns especially systems in which organic molecules are present.

2.1.2 Periodic boundary conditions

Periodic boundary conditions (PBC) is a technique able to minimize the edge effects in a finite system. It consists in putting the molecules to be simulated into a space-filling box and in surrounding it by periodic copies of itself in all directions. The atoms in the periodic images are not subjected to the simulation, but they get the motions from the main box through symmetry operations.

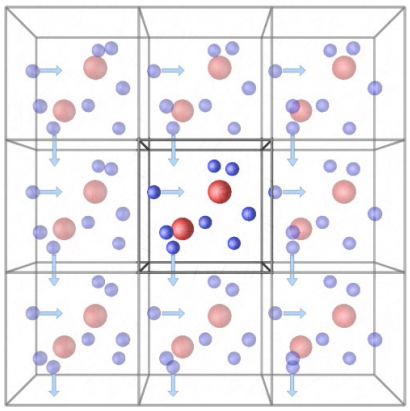


Figure 19 Schematic representation of periodic boundary conditions. Adapted from: http://lammpstube.com/2019/10/30/periodic-boundary-conditions/

Often (as in GROMACS) the pbc are used in combination with the minimum image convention, that is only one image of each particle is considered for the calculation of non-bonded energy, and this is the nearest one. Therefore, the cut-off radius for non-bonded interactions should be less than half the shortest box vector.

2.1.3 The solvation

The solvation influences the physico-chemical properties of the biological macromolecules and consequently its structure and dynamics. Among the main effects, it is worth citing the shielding of electrostatic forces and the hydrogen bonds formation between water and polar atoms of the solute. For these reasons, the presence of solvent in MD simulations have to be considered. The treatment of solvent can be implicit or explicit. A solvent is implicit when the water atoms are not

explicitly added in the system and their contribution is approximated with a continuum medium. One of the most used is the Generalized Born / Surface area (GBSA) model. However, implicit solvent methods are less accurate and the *ensamble* of conformations is not correctly sampled [47] [48]. In the explicit methods, solvent molecules are explicitly included and inserted in the box system. Several models for parametrization of explicit solvent molecules are available, among them SPC [49], SPC/E [50], TIP3P, TIP4P [51] and TIP5P [52], the choice depending on the used force field.

2.1.4 The energy minimization

To each molecular conformation is associated a potential energy from which we can derive the force acting on i^{th} atom, that is:

$$\vec{F_i} = -\frac{dV}{d\vec{r_i}} \tag{3}$$

In an atomic system, if the total force is zero, the system is in equilibrium and the potential energy is in a minimum. Energy minimization is a procedure that allows us to reach the relative minimum of a system, which corresponds to an energetically stable state. Before performing a molecular dynamics simulation, an energy minimization is required, to avoid structural distortions in the molecules due to the presence of unbalanced forces.

Several algorithms are used to perform a minimization. Some of them are more useful when the starting conformation has an energy far from the minimum (e.g. steepest descent, simplex...), others are used to optimize conformations close to the minimum (e.g. Newton-Raphson, conjugate gradients…). The best is to alternate them to reach the desired potential energy gradient.

2.1.5 The MD trajectory

A scheme to summarize the steps that permit to calculate the trajectory of each atoms of the system simultaneously, and therefore to monitor the conformational changes of the macromolecule, is below proposed:

MD TRAJECTORY CALCULATION

1) Initial conditions

- Initial positions
- Initial velocities
- Potential Energy

⇩

2) Compute forces on any atom

$$F_i = -\frac{\delta V}{\delta r_i}$$

⇩

3) Integration of motion equations

$$\frac{dv_i}{dt} = \frac{F_i}{m_i}$$

and

$$\frac{dr_i}{dt} = v_i$$

⇩

4) Writing the new positions and velocities

The 2, 3, 4 points are repeated for the required number of integration steps. The integration algorithms have to be time-efficient and conserve energy.

The default algorithm for the integration of motion equations, in GROMACS, is the *Leap-Frog* algorithm [53]. It consists on computing the positions at time t and the velocities at time t - $\frac{1}{2}\Delta t$, using these equations:

$$\boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) = \boldsymbol{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}\boldsymbol{F}(t) \quad (4)$$

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t\boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) \quad (5)$$

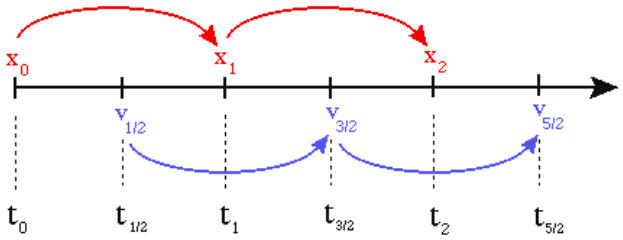


Figure 20 The leap-frog integrator method

The iteration of these operations allows to obtain the trajectory of the system. The choice of the time step Δt is important: it should be enough small to be accurate and to avoid meaningful deviations from the correct analytical trajectory, but it should be also as big as possible to not excessively overload the computational time. The most largely used integration algorithms are robust enough to allow the use of a Δt of 1 or 2 fs.

The initial positions are defined by three-dimensional structures modelled or resolved by X ray crystal diffraction or NMR. The initial velocities are randomly assigned to atoms, following the Maxwell-Boltzmann distribution, in dependence from the desired temperature T.

During the simulation, all velocities are rescaled in order to keep the temperature constant.

2.1.6 Standard simulation protocol

A brief protocol of a standard simulation of a protein system is summarized below.

1) Choice of the starting structure and its minimization
2) Solvation, with addition of:
    a. Water
    b. Ions (to neutralize the system or/and to reach physiological ionic force conditions)
3) Energy minimization of the whole system
4) Position restrained molecular dynamics: only solvent motions are allowed, leaving it to relax around the protein
5) Full motion molecular dynamics: the whole system is free to move. It includes:
    a. An equilibration phase, in which the system reaches the equilibrium of macroscopic parameters (e.g. temperature, pressure, energy). It often includes a preceding heating phase, in which the temperature is slowly increased to the fixed value.

b. A production phase, in which the real full MD is performed and the trajectory is generated.

2.1.7 Limits of MD technique

The major limitation of the molecular dynamics simulations regards both the system complexity, which forces the use of approximations in the interaction functions, and the system dimensions, that could greatly increase the computational cost, leading to a poor conformational sampling. Some solutions are possible to accelerate the calculation. Metadynamics, replica exchange dynamics and steered molecular dynamics, are some of the methods used to overcome these limits, depending on the system properties under study. Another solution is to reduce degrees of freedom by using Coarse-grained force-fields (CG). A second limitation regards the classical nature of Newton equations: a classical MD can not simulate processes like chemical or enzymatic reactions, that involve quantum effects, unless hybrid quantum mechanical/molecular mechanical methods are used.

3. Experimental techniques

Spectroscopic methods are widely implemented to provide information about molecular kinetics or molecular structure [54]. For the aim of this work, UV/visible spectroscopy and FTIR spectroscopy have been used to study the collagen self-assembly process and the collagen dehydration-rehydration cycle, respectively. In this chapter we provide a description of these two methods.

3.1 UV/visible spectroscopy

The UV/visible spectroscopy permits studies on kinetics of reaction, protein folding, dynamics and interactions. It involves the temporary exchange of energy between light and matter, that becomes excited. The absorption of UV/visible light is allowed by electronic transitions, in fact if the light that hits a chromophore has got enough energy to produce an electronic transition, the matter absorbs photons and electrons move on to an orbital with higher energy. The excited electrons will release their energy returning to their ground states using possible paths like heat dissipation, fluorescence, phosphorescence, energy transfer, FRET, photochemical reaction [54].

The visible light region comprises wavelength in the range 780-380 nm, while the UV region is divided into:

- Near UV, 380-200 nm
- Far UV, 200-100 nm

The absorption of UV light by proteins involves energy levels of

- Peptide bonds, which strongly absorb at 190 nm and weakly absorb at 210-220 nm.
- Sidechains of aromatic residues, which have transitions in the near-UV. The absorption bands for Phe is at 257 nm, for Tyr is at 274 nm and for Trp is at 280 nm [54].

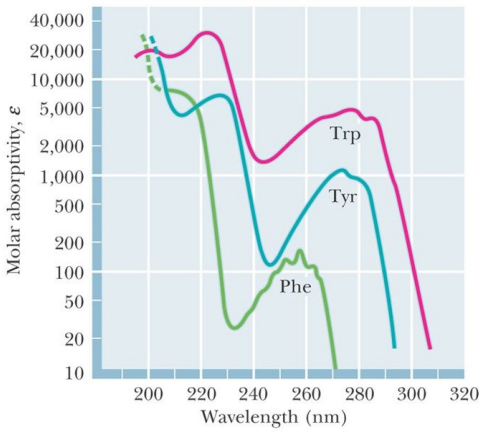


Figure 21. The UV spectra of aromatic residues at pH 6. Adapted from [55]

A spectrophotometer records the intensity of incident light I_0 and transmission light *I*. The ratio of these two intensities is called transmittance, T

$$T = \frac{I}{I_0} \tag{6}$$

and indicates the fraction of incident light that is transmitted. The absorbance, instead, indicates the quantity of absorbed light and it is expressed as:

$$A = \log(\frac{1}{T}) \tag{7}$$

One of the most common applications of UV/visible absorption spectroscopy is to detect the concentration C of a species in solution using the Beer-Lambert law:

$$A = \varepsilon \mathrm{Cl} \tag{8}$$

where l is the optical path length and ε is the molar extinction coefficient [56].

3.1.1 Spectral Artifacts: light scattering sample

Many macromolecular species are of comparable size to the wavelength of light, and as a result they scatter light [56]. Protein aggregates formed in aqueous solution can reach micrometer size range and so they scatter light in the UV-visible region. This phenomenon causes turbidity of solution, meaning a loss in intensity of transmitted light, and it can be a useful method for following protein aggregation [57]. Considering that the aim of our study is to monitor the collagen self-assembly, we recorded the changes in the absorbance of collagen solution, caused by the increase

of the turbidity of solution when the collagen molecules aggregated (figure 22). For this reason, the spectrophotometer was set at fixed wavelength mode at 310 nm, and growing absorbance values recorded were not caused by electronic transitions but by a growing reduction of transmittance, due to transition from a transparent solution to a turbid solution.

Figure 22. Examples of turbid collagen solutions: the collagen self-assembly increases the turbidity of solution.

3.1.1 UV/visible spectrophotometer

A spectrophotometer is composed by a light source, a sample chamber, a monochromator or a prism, and a detector. The main radiation sources are a Tungsten filament (300–2500 nm), a deuterium arc lamp (190–400 nm), Xenon arc lamp (160–2nm). Typical detectors are photomultiplier tubes, photodiodes, a photodiode array or a charge-coupled device (CCD). Scanning monochromators are used to filter a single wavelength.

Spectrophotometers can be either single beam or double beam. In the first kind of instrument all the light passes through the sample cell. To obtain blank measurements the cuvette containing only a solvent has to be recorded first. This type of instrument is the earliest design and is still commonly used in both teaching and industrial labs. In a double-beam spectrophotometer, the light, before reaching the sample, is split into two beams. One beam passes through the reference cell and the other one passes through the sample. The measurements are recorded as the ratio of the two beam intensities. Some instruments have two detectors to measure reference and sample at the same time. Other ones have a beam chopper, through which the two beams pass and which blocks one at a time. In this case, the detector

alternates the measure of the reference beam and of the sample beam in synchronism with the chopper.

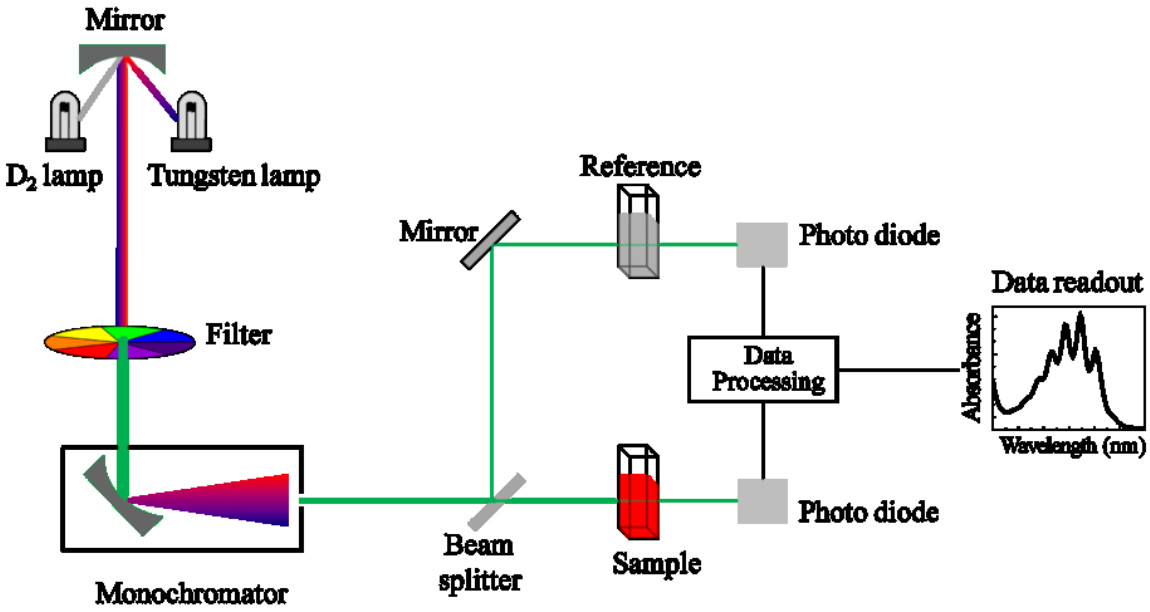


Figure 23. Schematic representation of UV/Visible double beam spectrophotometer, from: http://www.wikipedia.org

The most of samples used for UV/Visible spectrophotometry are liquids and are inserted in transparent cells, known as a cuvettes. A typical cuvette is made of quartz glass, transparent throughout the UV, visible and near infrared regions.

3.2 IR spectroscopy

The infrared spectroscopy is a spectroscopic technique commonly used to identify organic and inorganic materials. The infrared wavelength region covers a range from 0.78 µm to 1000 µm of the electromagnetic spectrum and is divided into:

- Near IR, 0.78–2.5 µm
- Medium IR, 2.5–25 µm
- Far IR, 25–1000 µm

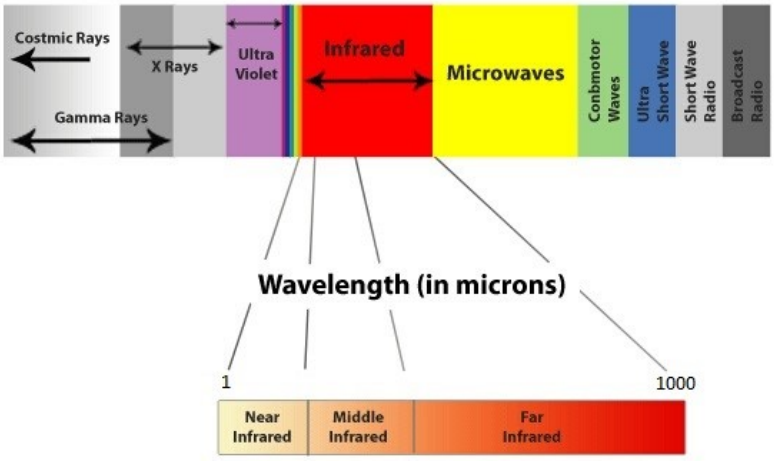


Figure 24. Electromagnetic spectrum with a focus on the IR portion

The absorption of IR light by the matter causes vibrational and rotational energetic transitions.

3.2.1 Molecular vibrations

The nuclei in a molecule vibrate around their equilibrium positions. These collective movements may be described by means of the model of the simple harmonic oscillator, consisting of two nuclei connected by a bond with force constant *f*. The vibrational frequency of the oscillator is given by:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{f}{m_r}} \tag{9}$$

where f is the constant force between two nuclei and m_r is the reduced mass ($1/m_r = 1/m_1 + 1/m_2$) [56]. The molecular vibrations can be divided into two main categories: *stretching* and *bending*. In the stretching mode the bond length changes, whereas in the bending mode the bond angle changes. The stretching mode can be symmetric or antisymmetric, instead scissoring, twisting, wagging, rocking are types of bending modes [58].

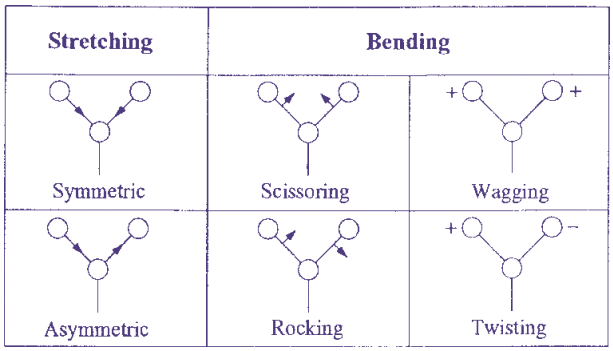


Figure 25. Scheme of stretching and bending vibrational modes. Adapted from [59]

3.2.2 The selection rule

The section rule claims that the vibrations are IR active if they cause a variation of dipole moment.

From quantum theory, it is possible to describe the molecular state as an eigenstate ψ, function of electrons and nuclei positions. According to the Born-Oppenheimer approximation, electronic and nuclear motions may be separated, and a pure vibrational transition can be expressed in form:

$$\Psi_0(\boldsymbol{r},\boldsymbol{R})\varphi_v(\boldsymbol{R}) \rightarrow \Psi_0(\boldsymbol{r},\boldsymbol{R})\varphi_{v'}(\boldsymbol{R}) \tag{10}$$

where $\varphi_v(\boldsymbol{R})$ and $\varphi_{v'}(\boldsymbol{R})$ are two vibrational states, R and r are the spatial coordinates of nuclei and electrons, respectively, and $\Psi_0(\boldsymbol{r},\boldsymbol{R})$ is the electronic eigenstate of the ground state.

When a molecule is hit by light, the electronic charge distribution is modified by the presence of the oscillating electric field,

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{i\omega t} \tag{11}$$

The molecular charge distribution is expressed as a series expansion of multipole, where, for neutral molecules, the first term is the dipole moment

$$\boldsymbol{\mu} = \sum_i e_i \boldsymbol{r}_i \tag{12}$$

where e_i is the electron charge and r_i is its position.

We can express the perturbation induced by vibrational transition through the potential:

$$V(t) = \boldsymbol{\mu} \cdot E_0 e^{i\omega t} \tag{13}$$

The transition dipole moment results from perturbative theory and for a vibrational transition it corresponds to:

$$< \Psi_0\varphi_\nu \mid \boldsymbol{\mu} \mid \Psi_0\varphi_{\nu'} > \tag{14}$$

If the nuclei are static, it is zero, whereas if they vibrate, the operator $\boldsymbol{\mu}$ depends on $\boldsymbol{R}$, because the electrons distribution changes as the nuclei positions change. The dipole moment $\boldsymbol{\mu}$ can be written as Taylor series expansion on the equilibrium position $\boldsymbol{R_0}$

$$\boldsymbol{\mu}(\boldsymbol{r},\boldsymbol{R}) = \boldsymbol{\mu}(\boldsymbol{r},\boldsymbol{R}_0) + (\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{R}})_0(\boldsymbol{R}-\boldsymbol{R}_0)+\ldots \tag{15}$$

The expectation value of the first term is zero, whereas the second one is not zero, but:

$$< \Psi_0\varphi_\nu \mid \boldsymbol{\mu} \mid \Psi_0\varphi_{\nu'} > = < \Psi_0 \mid (\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{R}})_0 \mid \Psi_0 >< \varphi_\nu \mid \boldsymbol{R} \mid \varphi_{\nu'} > \tag{16}$$

The first term is called electronic integral, whereas the second one is vibrational integral. The electronic integral describes the change in the permanent dipole of the electronic ground state with the change of the atomic positions in a molecule. When the dipole moment does not change, its derivative, as also the transition dipole moment, is null, hence the vibrational transition does not occur.

The selection rule, hence, suggests that not all the vibrations produce dipole moment variation. The vibrations that are able to change the dipole moment are called IR active. For example, the dipole moment of CO_2 molecule is null when the resulting vibration is the symmetric stretching mode, whereas it changes during asymmetric stretching and bending mode.

3.2.3 FTIR Spectroscopy

The Fourier Transform infrared spectroscopy is the most widely used technique to acquire IR spectra, because it allows to simultaneously measure the intensity of a range of frequencies. For this reason, this method speeds up the acquisition process compared to IR spectroscopy. The term “Fourier Transform” refers to the conversion of intensity data in the resulting spectra by means of Fourier Transform [3].

3.2.3.1 FTIR spectrophotometer

A FTIR spectrophotometer is composed by the infrared source, the sample chamber, the interferometer and the detector.

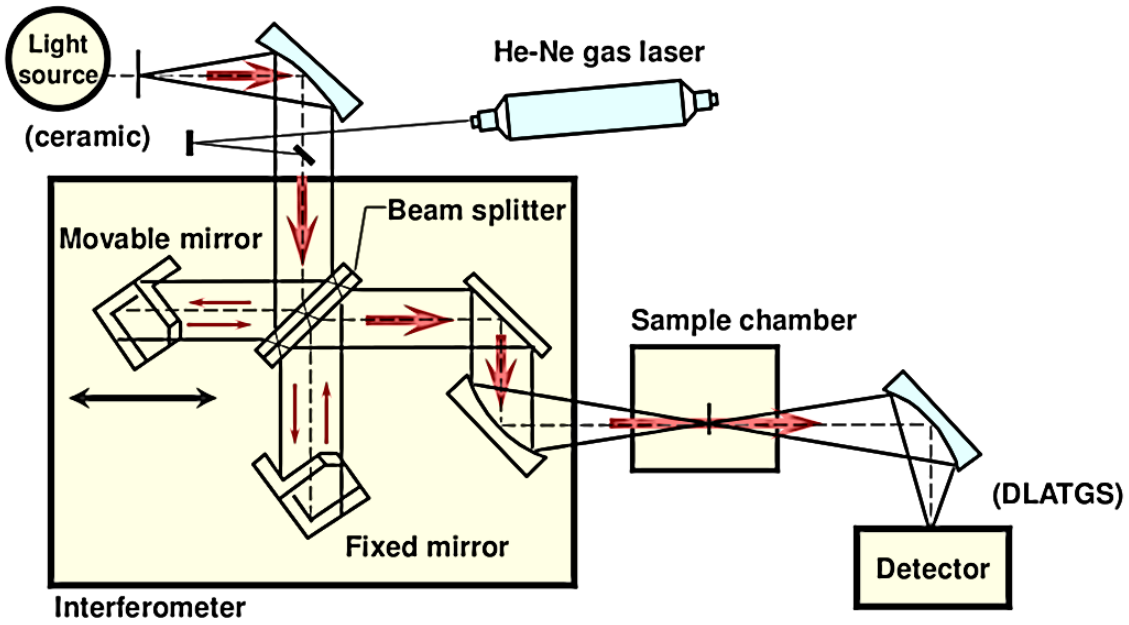


Figure 26. Schematic representation of FTIR spectrophotometer. Adapted from Covalent Metrology Services (https://covalentmetrology.com/ftir/)

The interferometer is kept at nitrogen flow in order to prevent the lenses from fogging up.

The radiation emitted by the source, entering in the interferometer, is collected by a beam split, which divides it into two rays of half intensity with regard to the original beam. These two beams reach two mirrors, one fixed and the other one movable in order to vary the optical path. The two beams reflected by the mirrors are combined achieving an interferogram that, then, is converted into resulting spectrum by means of Fourier Transform [3].

The interferogram intensity is:

$$I_{interferogram}(t) = k\int_{-\infty}^{+\infty} I_{beam}(v)e^{i2\pi vt}dv \qquad (17)$$

where I_{beam} is the intensity of the radiation, v is the wave number and k is a constant.

The beam intensity is calculated from the interferogram intensity by inverse Fourier Transform:

$$I_{beam}(\nu) = k \int_{-\infty}^{+\infty} I_{interferogram}(t) e^{-i2\pi\nu t} dt \quad (18)$$

From beam intensity the absorption spectrum is generated calculating the logarithm of the intensity quotient of blank to sample. Blank measurements are obtained by recording the interferogram with the empty sample cell (without sample) [3].

3.2.3.2 Application of FTIR spectroscopy to biological systems

FTIR spectroscopy is a useful method to study biological macromolecules and to obtain structural information on them. In particular, in the last years, it was employed to investigate the hydration events of biopolymers [60] [61]. As regards the study of hydration proprieties, it enables to measure the position and amplitude changes of the absorption bands concerning the functional macromolecules groups, as a function of water content, gaining information on the water molecules decorating the biomolecules [3].

Vibrational modes of water can be detected by using FTIR spectrophotometer, in fact the IR spectrum of water displays three main bands corresponding to three different vibrational modes: OH stretching, bending and libration modes. The contribution of these bands is displayed in the spectra of hydrated macromolecules and the change of hydration conditions of the sample causes changes in the IR spectral pattern. For example, the OH stretching band (3600-3200 cm^{-1}) can be analyzed in order to obtain information on the water sorption process during a dehydration/rehydration cycle of a sample [3]. Water adsorption-desorption isotherms can be obtained from a mathematical analysis of this band. The shape of OH stretching band depends on the length of OH bond: changes in the OH bond length cause changes in the H bonds network of water molecules. These alterations induce spectral variations, like shifts of peak position or variations of band amplitude. The OH stretching band analysis is performed until after the subtraction of "dry sample" spectrum, i.e. the spectrum of sample lacking of water molecules as a result of heating to 80°C. This prior operation is carried out in order to remove any protein contribution to the band of C–H and N–H stretching vibrations, which occur in the same wave numbers range [3].

As regards the study of protein samples, another useful analysis of IR spectra concerns Amide regions. Two main bands, called Amide I (1650 cm^{-1}) and Amide II (1550 cm^{-1}), correspond to vibrational modes of peptide bond: the stretching of

peptidic carbonyl group is responsible for the appearance of Amide I band, whereas the stretching of amidic bond and deformation of C-N-H are responsible for the formation of Amide II band. Amide I band is analyzed in order to gain information on the type of secondary structure forming the protein, in fact band position and shape critically depend on the secondary structure of the protein.

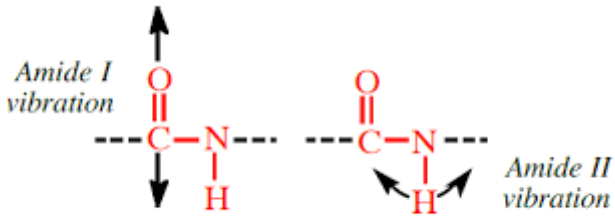


Figure 27. Vibrational modes of peptide bond

Peptidic bond generates 7 bands in the infrared region: in addition to Amide I and II, other known Amide bands are Amide A, caused by the stretching of the amino group and Amide III, which depends on the lateral chain motions and hydrogen bonds.

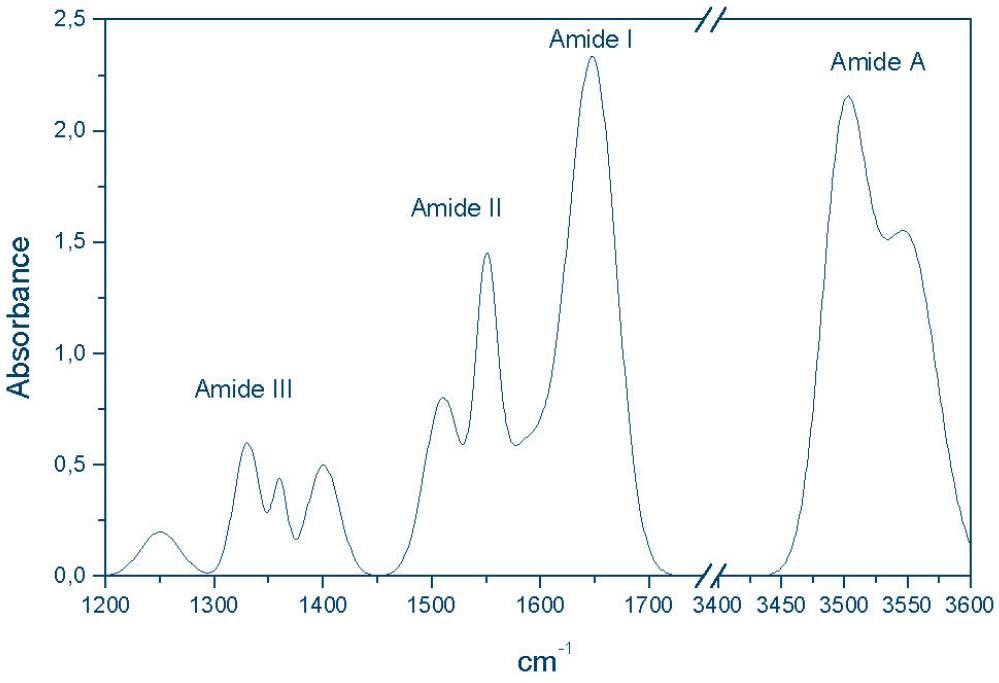


Figure 28. Amide bands of IR spectrum, adapted from JenaLib [62]

The picture above shows the four main Amide bands. Actually, the proteins cause up to seven bands: the three not cited bands have intensities so weak that it is not possible to use them for protein structural analysis.

4.Materials and methods

Rat tail type I collagen was chosen to perform experimental measurements due to its easy availability. Consequently, also MD simulations were performed on the same system, allowing the comparison of results.

4.1 Molecular dynamics simulations

In this section, the model building of collagen molecules and the set-up of MD simulations are described. The paragraph "Molecular Dynamics simulations set-up" is split in two sub-paragraphs, referring to the aggregation study and to the dehydration study.

4.1.1 Model building of triple helices

For lack of suitable solved structure in the PDB database [63], we had to build a model for tropocollagen, by using α1 and α2 chain sequences of *rattus norvegicus* type I collagen, available in the Uniprot database [64] (entry: P02454 and P02466). These sequences are about a thousand residues long, so too big to perform an atomistic molecular dynamics study. For this reason, we built short tropocollagen fragments, whose sequences had been selected from the whole rat collagen sequence. In particular, we decided to separately simulate the aggregation of two kind of triple helices with different hydropathic profiles, to determine the influence of amino acid composition on the aggregation process. Because the main difference between the two selected fragments is the charges content, we named them "lowly charged" (LC) and "highly charged" (HC) tropocollagens. Thereafter, we selected other fragments from the same collagen sequence to obtain 234 residues staggered tropocollagens. In fact, these sequences had been chosen with a shift of m x 234 residues (with m = 1 or 2) downstream and upstream of both LC and HC models. The hydropathic features of staggered sequences match. Henceforth, we refer to lowly charged fragment and its staggered sequences as "LC model", and to highly charged fragment and its staggered sequences as "HC model".

Tropocollagen of LC model	α1 chain	α2 chain
T0	A**P**GIAGA**P**GF**P**GARGPSGPQGPSGA**P**G PKG	L**P**GVAGA**P**GL**P**GPRGI**P**GPVGAAGATG PRG
T1	PAGS**P**GFQGL**P**GPAGP**P**GEAGK**P**GEQG V**P**G	PAGP**P**GFQGL**P**GPSGTAGEVGK**P**GERG LPG
T2	P**P**GATGF**P**GAAGRVG**P**PGPSGNAGP**P**G P**P**G	PPGMTGF**P**GAAGRTG**P**PGPSGITGP**P**G P**P**G
T3	DRGI**K**GHRGFSGLQGP**P**GS**P**GS**P**GEQG PSG	ARGLPGLKGHNGLQGLPGLAGLHGDQG APG

Tropocollagen of HC model	α1 chain	α2 chain
T0	M**K**GHRGFSGLDGAKGDTGPAGPKGE**P**G S**P**G	F**K**GIRGHNGLDGLKGQ**P**GAQGV**K**GE**P**G A**P**G
T1	ADGVAGPKGPAGERGS**P**GPAGPKGS**P**G EAG	ADGRAGVMGPPGNRGSTGPAGVRGPNG DAG
T2	ERGAAGL**P**GPKGDRGDAGPKGADGSPG KDG	ERGAAGI**P**GGKGEKGETGLRGEIGNPG RDG
T3	QRGERGF**P**GL**P**GPSGE**P**GKQGPSGASG ERG	SRGERGQ**P**GIAGALGE**P**GPLGIAGP**P**G ARG

Table 1 Sequences of shifted tropocollagens in LC and HC models. The bold residues are hydroxylated. Adapted from [65].

Residue type	T0	T1	T2	T3
Hydrophobic	71.6%	63.6%	69.6%	57%
Polar not charged	21.6%	25%	27%	28%
Charged	6.8%	11.4%	3.4%	15%
HYP	13.6%	15%	16%	6.7%

Residue type	T0	T1	T2	T3
Hydrophobic	56.7%	68%	59%	55.5%
Polar not charged	21.1%	13%	8%	23.4%
Charged	22.2%	19%	33%	21.1%
HYP	7.7%	4.4%	3.3%	10%

Table 2 Percentages of residue type in LC and HC tropocollagens. Adapted from [65]**.**

Actually, the rat tail collagen structure is included in the PDB database [63] (id code: 3HQV [13]), but it consists of C_α atoms only, at very low resolution (5.16 Å), so not appropriate to perform MD simulations. Therefore, in order to build first raw models, we used the following PDB templates: 2KLW [66] containing several charged residues, suitable for every fragment of HC model, and 1WZB [67], rich of POG triplets, suitable for every fragment of LC model. Swiss-model server [68] was used to build tropocollagens, each chain of which is 29 or 30 residues long, depending on the template length.

In order to complete our models building, we used Vienna PTM 2.0 server [69] [70] to mutate some selected lysines and prolines in hydroxylysines and hydroxyprolines, as specify in the 3HQV structure [13].

We built four tropocollagens for each model, in order to obtain a microfibril after their aggregation. In fact, a microfibril is known to be composed by four or five staggered tropocollagens [10]. However, due to the shift of 234 residues, the fifth segment should comprehend the unordered telopeptide region, so we used only four fragments composed by only Gly-Xaa-Yaa triplets in triple helix structure.

4.1.2 Molecular dynamics simulations set-up

All molecular dynamics simulations were performed by means of GROMACS software package [38], with the GromosG54a7 force-field [71]. In every simulation, SPC/E solvent model was used and N- and C-termini of the tropocollagen fragments were maintained in the neutral form to take into account the presence of the rest of the sequence. The calculation of Van der Waals interactions and electrostatic contribution was implemented by means of potential switch and PME summation

methods, respectively. Before each full MD simulation, firstly, we carried out a minimization of the whole system, secondly, we performed a 100 ps position restrained MD simulation to relax the solvent around solute molecules. All the simulations (position restrained and full) were performed with an integration step of 2 fs.

4.1.2.1 MD simulations set-up details for the study of collagen self-assembly

In order to study collagen self-assembly, several MD simulations were performed both on LC and HC model systems. These systems are composed by a number of tropocollagens from two to four. In each system, the tropocollagens were embedded in a cubic box of water (11x11x11 nm for HC model and 10x10x10 nm for LC model) with periodic boundary conditions, and, initially, placed at 2.0 nm distance between each other using VMD software [72]. For each simulation, we selected an NTP ensemble, with V-rescale thermostat for temperature coupling and Berendsen barostat for pressure coupling. The temperature was fixed at 300 K, and the pressure was kept at 1 atm. We studied the aggregation process in two conditions: at physiological salt concentration and at low ionic strength.

Simulations at physiological salt concentration

In these conditions, we investigated the aggregation process of two, three and four tropocollagens, hence we had prepared and then simulated three systems for both LC and HC models. In order to neutralize the systems and reach a physiological concentration (0.1 M), Na and Cl ions were added. We carried out full MD simulations 200 or 300 ns long (according to the RMSD profile, that checks the attained stability, see the 4.1.3 sub-section), each one repeated at least three times.

Simulations at low ionic strength

We studied the aggregation of three and four tropocollagens for both models at low ionic strength. Simulations at low ionic strength were obtained, adding only Cl ions enough to neutralize the system. Two replicas 200 ns long for each simulation were carried out.

4.1.2.2 MD simulations set-up for collagen dehydration study

The four tropocollagens microfibrils obtained from the previous simulations at physiological conditions have been the starting point of simulations in dehydrated state, and the last conformations of these ones were used in rehydrated state

simulations. We performed full MD simulations 0.5 or 0.4 μs long (according to the RMSD profile, that checks the attained stability, see the 4.1.3 sub-section). In the next two sub-paragraphs are described the systems and set-up details related to simulations in dehydrated state and in rehydrated state.

Simulations in dehydrated state

The microfibril models of LC and HC model, obtained by the aggregation of four tropocollagens in the previous simulations at physiological conditions, were used as starting structures for simulations in dehydrated conditions. A 3 Å layer of water, corresponding to the first hydration shell, was cut out around the microfibril models in the original water box. 10 and 8 Cl ions were added in the LC and HC systems, respectively, to neutralize them. In this way, the hydration values of LC and HC models are h = 0.42 (g water /g protein) and h = 0.49 (g water /g protein), respectively, corresponding to dry states. Indeed, it has been reported that the transition from the dry state to the solution state take place at h = 0.9-1.0 (g water /g protein) [73]. NVT ensemble was selected, to avoid molecules spreading all around the box. Pressure, however, was checked and it remains stable and does not assume any negative values. The box volume was 11x11x11 nm^3 for HC model and 10x10x10 nm^3 for LC model; T = 300 K, with V-rescale thermostat for temperature coupling.

Simulations in rehydrated state

Simulations in rehydrated state were performed adding water molecules to the final LC and HC models systems (i.e. collagen aggregates plus their water layers and Cl ions), obtained by the previous simulations in dehydrated conditions. Water molecules were added up to fill the box volumes. Afterwards, Na and other Cl ions were inserted to reach a physiological concentration (0.1 M), as the simulations at physiological conditions described in the 4.1.2.1 sub-paragraph. NTP ensamble was selected, where T = 300 K and P = 1 atm, with V-rescale thermostat for temperature coupling and Berendsen barostat for pressure coupling.

4.1.3 Methods for the MD trajectories analysis

The following programs were used to analyse MD trajectories:

- GROMACS subroutines [38]
- VMD software package [72]
- Conan tool [74]

- MD analysis software package [75] [76]
- Origin 2018 software package (OriginLab Corporation, www.originlab.com)

First of all, to verify the reached stability of a structure during a simulation, the Root Mean Square Deviation (RMSD) was calculated, by means of VMD software [72]. The RMSD indicates the differences between two structure calculating the averaged distance between atoms of superimposed proteins:

$$RMSD = \sqrt{\frac{1}{N}\sum(r_{i,a} - r_{i,b})^2} \tag{19}$$

where N is the number of superimposed atoms, r is the atom position, the subscript i is the index for i^{th} atom, and the subscripts a and b are referred to the two superimposed structures. In our simulations we evaluated the changes in the RMSD during the trajectory, after superimposing the structure of each frame to the starting structure.

4.1.3.1 Methods for the analysis of MD trajectories: a collagen self-assembly study

To follow the aggregation process, the distance between two tropocollagen fragments was monitored during the trajectory. The distance was calculated by a GROMACS subroutine [38], considering the center of mass (COM) of two triple helices.

Interactions and contacts were checked by two kind of analyses: calculation of inter-tropocollagen hydrogen bonds (with a 3.5 Å cut-off distance and 30° cut-off angle) and contact/interaction maps evaluation. The former was carried out by a GROMACS subroutine [38], while the latter was obtained by using Conan tool [74]. A contact map is a 2D distance plot, in which the distance between two amino acids is represented. This distance is defined as the shortest distance between any two heavy atoms of two different residues. Interaction map is a 2D graph, in which the type of interactions (hydrophobic, polar and electrostatic) between two contacting residues is shown. In our case, we chose to calculate the kind of interactions in the frame in which the first stable contacts arose, considering a minimum of fifteen points in the map.

4.1.3.2 Methods for the analysis of MD trajectories: a collagen dehydration study

In order to monitor structural changes induced in the assembly by dehydration, protein-protein hydrogen bonds and water-protein hydrogen bonds (with a 3.5 Å cut-off distance and 30° cut-off angle) were determined by a GROMACS subroutine [38]. Differential contact maps, that represent the difference with respect to the initial contact map, are calculated by using Conan tool [74]. Radius of gyration of the

collagen assembly around z axis was calculated as a function of time. The calculation, which considers only alpha carbon atoms, was performed by means of a GROMACS subroutine [38], using the following formula:

$$R_{g,z} = \sqrt{\frac{\sum_i (r_{i,x}^2 + r_{i,y}^2) m_i}{\sum_i m_i}} \tag{20}$$

where m_i is the mass of i^{th} alpha carbon atom, and $r_{i,x}$, $r_{i,y}$ are the components on x and y axes of its position vector $\boldsymbol{r}$ with respect to the center of mass (COM) of the protein assembly.

We investigated also the changes in water behavior, induced by dehydration and the next rehydration, checking the water occupancy and persistence. The former analysis was carried out by the building of volumetric maps in VMD software package [72], while the latter was studied by means of MD Analysis tool [75] [74]. In particular, the persistence of water molecules that lie within 2 Å of protein was determined by calculating their survival probability SP(t), which measures the average number of water molecules that survive around the protein after a time t. Therefore, fraction of water molecules lying within 2 Å layer at time t' and persisting after a time t'+t, was calculated by the following formula [77]:

$$SP(t) = \frac{1}{T} \sum_{t}^{T} \frac{N(t', t)}{N(t')} \tag{21}$$

where N(t') is the number of waters that lie within the selected layer at time t', N(t',t) is the number of water molecules that persists within the selection at every frame from t' to t' + t, and T is the number of frame steps considered in the $SP(t)$ calculation. The curve fitting of SP(t) function allows us to extract a useful parameter called "relaxation time" or "residence time". Different theories about which is the best function fitting the survival probability curve are described in literature [78]. In some cases a single exponential function is chosen, extracting a single residence time [78] [79] [80] [81], in other cases a combination of exponentials is preferred, obtaining more than one relaxation times [82] [83]. In particular, stretched exponential was reported to be better with water relaxation time less than 15 ps (fast relaxation), whereas an exponential decay is used with times between 100 and 200 ps, related to the slow component of water relaxation [83]. In agreement with literature, we chose to fit SP(t) with a single exponential function, being our time steps of 50 ps.

4.2 Experimental studies

In this section the collagen samples preparation and measurements details, to study the self-assembly process and the dehydration-rehydration cycle, are described. In every experiment, lyophilized rat tail collagen, purchased by Sigma-Aldrich, was used.

4.2.1 UV-visible measurements to study collagen self-assembly

Collagen self-assembly kinetics was studied at three different pH values by means of UV-visible spectroscopy.

4.2.1.1 Materials

In order to prepare a collagen solution of 1mg/ml, the lyophilized collagen was dissolved in acetic acid 0.5 M. This solution was dialyzed, for 12 h and at 4 °C, against phosphate buffer 67mM at three different pH values. Finally, the three samples obtained at pH 8.0, 7.0 and 5.0 were centrifuged at 10000 *g* for 10 min, to remove any precipitate [84].

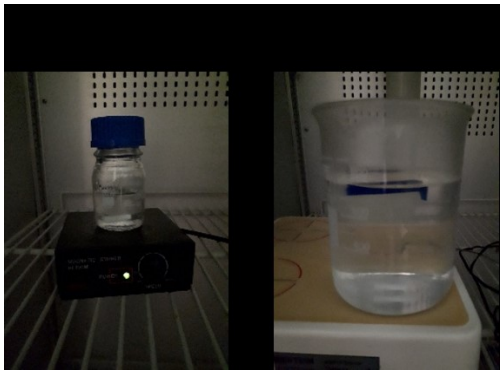


Figure 29. The first two steps of sample preparation protocol. a) Dissolution of collagen in acetic acid, b) dialysis of solution against phosphate buffer.

4.2.1.2 Measurements

We monitored the collagen self-assembly at three different pH values by recording the turbidity changes of the collagen solutions. According to the method of Nomura and co-workers [26], we measured the absorbance increase at 310 nm, as a function of time. The data were collected every 10 s at 300 K by UV-visible spectrophotometer Jasco 7850, equipped with a thermostat.

4.2.1.3 Data analysis

The kinetics of collagen self-assembly was analysed by means of Origin 2018 software package (OriginLab Corporation, www.originlab.com). The curves were plotted as aggregate fraction (AF) vs time and each one shows a sigmoidal profile, according to the literature [85]. AF was obtained by normalizing the absorbance data to the maximum absorbance. From the plotted curves we estimated the lag time and the half-time. The former was extrapolated from the intercept on the time axis of the tangent at t*, that is the time corresponding to the maximum value of the growth rate slope (V_{max}), as calculated by the derivative of the curve at the inflection point. The latter was estimated observing the time in which half of the self-assembly process took place, i.e. when half of fibrillar aggregates was formed.

4.2.2 FTIR measurements to study collagen dehydration-rehydration cycle

FTIR measurements allowed us to observe the changes in the water network, which solvates collagen, induced by a dehydration process, and if the event course is reversible after rehydration. In order to achieve a dehydration-rehydration cycle, we used saturated salts solutions to assess the atmosphere humidity able to induce the dehydration and the following rehydration in the collagen sample which was submitted to IR measurements. We also carried out a gravimetric study in order to determinate water content in a parallel collagen sample during the dehydration-rehydration cycle.

4.2.2.1 Materials

For infrared measurements a collagen aqueous solution (15 mg/ml) was prepared and deposited on a calcium fluoride window, obtaining a thin film allowed to dry in air under ambient conditions. Thereafter, this CaF_2 platelet, with the smeared sample, was assembled in a sealed sample support, that consists in a dry box equipped with IR transparent KrS_5 windows. The dry box incorporates a vertical CaF_2 platelet holder, allowing for transmission measurements, and a vessel, that contains the saturated salt solutions used to dehydrated-rehydrated the sample. The sealed box allows to isolate the sample from the external atmosphere and is inserted in the FTIR spectrophotometer sample chamber.

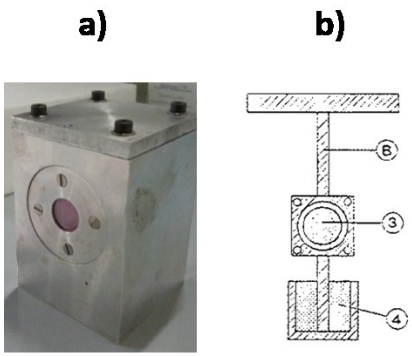


Figure 30. a) Dry box. b) The sample support inside the dry box: vertical platelet holder (3) and vessel (4).

4.2.2.2 Experimental procedures and measurements

Saturated salt solutions were used to obtain the desired relative humidity (RH) inside the box, in order to dehydrate or rehydrate the collagen sample. Some salt solutions, in saturation condition, are able to control the known water activities a_w:

$$a_w = \frac{p}{p_0} \tag{22}$$

where p is the water vapor of solution and p_0 is the pure water vapor pressure. Table 3 shows the salts used and the respective activity water values.

Salts	a_w
K_2SO_4	0.97
KNO_3	0.92
KCl	0.84
KBr	0.81
KI	0.69
$NaBr$	0.58
$Ca(NO_3)_2$	0.51
$CaCl_2$	0.29
$CaBr_2$	0.16
$LiCl$	0.11
$NaOH$	0.6

Table 3 Salt solutions employed to dehydrate/rehydrate the collagen sample and their corresponding activity water values.

The collagen film together with the salt solution, therefore, were left inside the dry box for two days, in order to equilibrate at the desired humidity level. After this time, the sample is ready to be submitted to the measurements, inserting directly the sealed dry box in the spectrophotometer, kept at N_2 atmosphere. The spectrum of CaF_2 window in the box was also recorded and subtracted from the spectra of CaF_2 platelet with the smeared film, in order to remove its contribution and to obtain only the contribution of the sample. The spectra were collected, firstly, by decreasing (dehydration run) and, secondly, by increasing (rehydration run) the relative humidity in the dry box. After the end of dehydration-rehydration cycle, the smeared film was heated to 80°C, keeping it in an oven for two hours, in order to obtain the "dry" sample and to record its spectrum. All spectra were collected by FTIR/IR spectrophotometer Jasco 240, operating at a temperature close to room temperature (290-300 K) in the transmission mode within the 4000-400 cm^{-1} range in 128 scans at spectral resolution of 2 cm^{-1}.

Concurrently, we carried out gravimetric measurements on 9 mg of lyophilized collagen, placed inside a dry box, which contained a vessel in which we insert the salt solutions to obtain a dehydration-rehydration cycle. Collagen was left inside the dry box for three days, in order to reach the opportune humidity level, and then it was weighed with an analytical balance (sensitivity of 10^{-4} g).

4.2.2.3 Spectra analysis

Two regions of the recorded spectra were analyzed by using of OriginLab program: OH-stretching and Amide regions.

The behavior of water molecules was monitored by studying of OH stretching band (4000-2500 cm^{-1}). Firstly, the spectrum of "dry" sample was subtracted from each spectrum recorded in the dehydration-rehydration process, in order to remove any protein contribution to the band of C–H and N–H stretching vibrations, which occur in the same wave numbers range. Secondly, the areas of the OH bands were calculated and plotted vs water activity (a_w), to graph the sorption water isotherms. This analysis is performed because the water surface coverage can be quantified by the integrated absorbance, assuming a modified form of Beer-Lambert law:

$$\tilde{A} = \int_{band} A d\nu = \varepsilon(\nu_{max}) \tilde{c} d \qquad (23)$$

where $\tilde{A}$ is the area of the band, $\varepsilon(\nu_{max})$ (L mol^{-1} cm^{-1}) is the molar absorptivity at maximum wave number corresponding to the peak of the band, $\tilde{c}$ (mol/L) is the concentration of the IR absorbing species and *d (cm)* is the average thickness of the IR absorbing species film, in this case adsorbed water [3]. It is important to bear in mind that $\tilde{A}$ is proportional to the concentration of molecules that produce IR active vibrations (in this case OH stretching) and not all water molecules can contribute to the signal. For example, the low fluctuations of the dipole moment of water molecules more fixed to protein surface could not be detected. The sorption water isotherms can be obtained also by gravimetric measurements, another way to quantify the variation of water content during the dehydration/rehydration cycle. The weight of dry sample was subtracted from weight of each sample subject to the dehydration-rehydration process, in order to determinate the weight of water. Finally, we plotted the h value (mg water/mg dry protein) as a function of a_w, with the aim to compare the water isotherms obtained from IR absorption spectra and from those gravimetric measurements.

Structural changes in the secondary structure of collagen induced by dehydration/rehydration process were estimated by the study of Amide I region. We operated the Amide band deconvolution for collagen before and after dehydration-rehydration cycle, by decomposition of the band in Gaussian sub bands, each one corresponding to a secondary structure, starting from a second derivative study.

5. Results and discussion: study of collagen self-assembly

In this section the main results, obtained from MD simulations and UV/Visible spectroscopy in the study of collagen self-assembly, are reported and explained.

5.1 MD simulations

MD simulations have been performed at physiological salt concentration and at low ionic strength for both LC and HC models, in order to discriminate the interactions involved in the aggregation processes.

5.1.1 Aggregation of LC fragments

In each simulation the LC tropocollagens were initially placed at 2 nm distance in a water box and their aggregation was tracked for 200 or 300 ns (according to the RMSD profile). Under physiological conditions (salt concentration of 0.1 M) the assembly of two, three and four triple helices was simulated and replicated three times, whereas at low ionic strength we simulated and repeated two times the aggregation of three and four tropocollagens.

5.1.1.1 Physiological salt concentration

In figure 31 the starting structures with the structures obtained at the end of the simulations are compared and the aggregation of two, three and four tropocollagens is shown

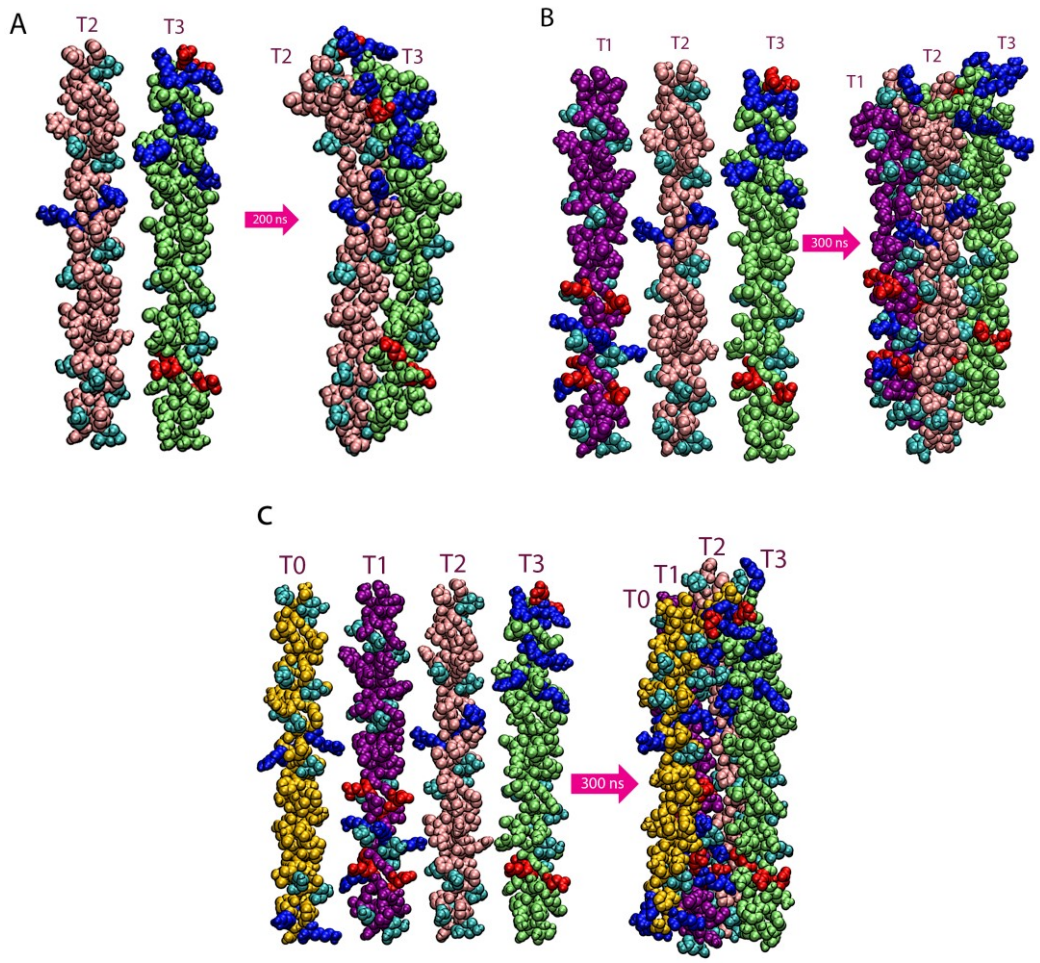


Figure 31. Aggregation of LC tropocollagens. Van der Waals representation of the systems with two (A), three (B) and four (C) fragments before and after 200 or 300 ns of MD simulations. Triple helices are labelled from T0 to T3 and colored in different way. Hyp and charged residues are highlighted with residue type coloring: Hyp is cyan, basic residues are blue and acidic ones are red. Adapted from [65]

The aggregation occurs in all three replicas of simulations with two and three triple helices, whereas the complete aggregation of four tropocollagens is observed only in two of four replicas, since in the other two the assembly is only partial. Contact maps of the last frame of the trajectory, reported in figure 32, help to visualize which regions of the tropocollagen chains become in contact.

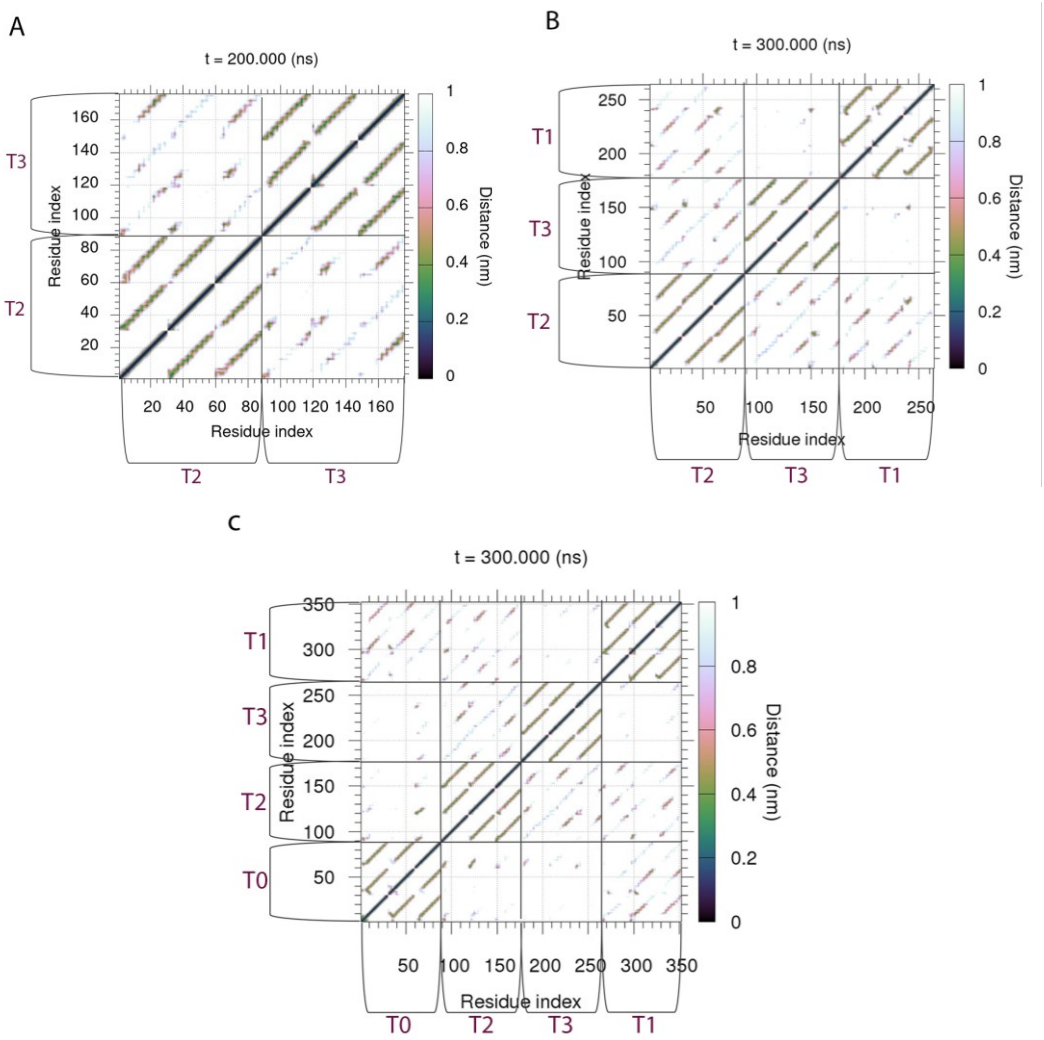


Figure 32. Contact maps of the last structure of the trajectories with two (A), three (B) and four (C) LC fragments. Adapted from [65]

Of particular interest is the arrangement of three and four tropocollagens: observing both figure 31 and figure 32, and comparing the regions in the maps with no contacts, it can be noticed that the aggregation of three triple helices is almost planar, whereas the assembly of four tropocollagens is nearly tubular. In fact, in figure 32B, the contacts between T1 and T3 triple helices are almost absent, confirming the planar arrangement of the three-fragments aggregate shown in Figure 31B. Conversely, the shape of the pseudo-fibril formed from four fragments shown in figure 31C, and the few contacts emerged between not adjacent tropocollagens, shown in figure 32C, reflect a more tubular assembly of the four triple helices.

To monitor the fragments approaching mechanism and investigate which are the kind of interactions that firstly drive the assembly, COM distances between

tropocollagen pairs were plotted as a function of time and interactions maps were also calculated in order to discover the type of contacts established between two triple helices at the first interactions (figure 33, 34 and 35).

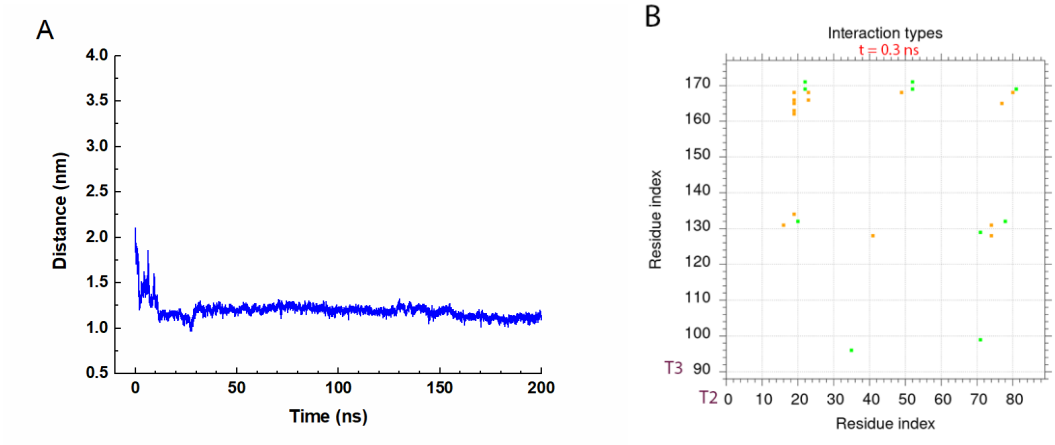


Figure 33. (A) Inter-tropocollagen COM distance during the two LC fragments (T2 and T3) trajectory. (B) Interactions map of first contacts in T2-T3 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

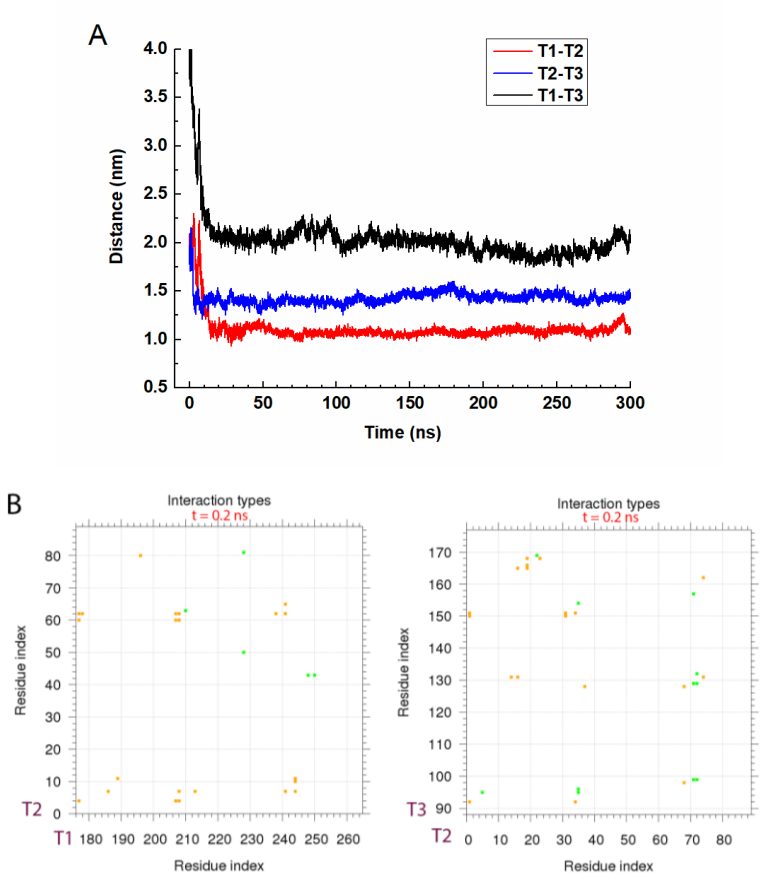


Figure 34. (A) Inter-tropocollagen COM distance during the three LC fragments (T1, T2 and T3) trajectory. (B) Interactions maps of first contacts in T2-T3 and T1-T2 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

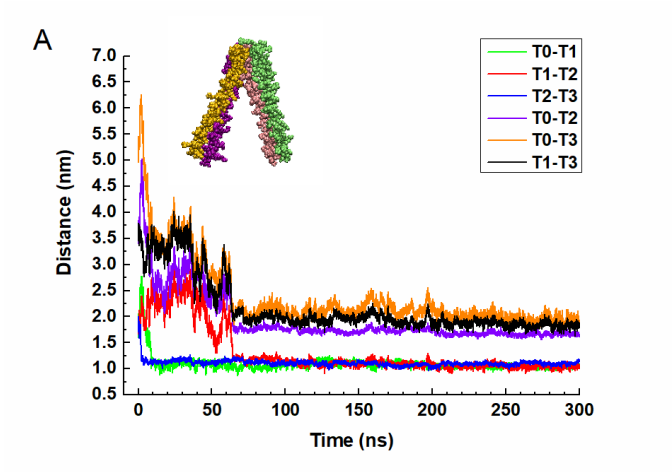


Figure 35. (A) Inter-tropocollagen COM distance during the four LC fragments trajectory. The inset image shows a representative frame after the first 55 ns indicating the formation of two pairs. (B) Interactions maps of first contacts in T0-T1, T1-T2 and T2-T3 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

As displayed in figure 33A and 34A, the approaching of fragments in simulations with two and three tropocollagens is very fast (<15 ns). The first interactions are mainly hydrophobic (orange spots). The aggregation of four tropocollagens, instead, is quite different, as observed in figure 35: the distance plot shows that not all the tropocollagens get close simultaneously, but two pairs (here T0-T1 and T2-T3) are formed first (as shown in the inset of figure 35A), and, only after 55 ns, the two dimers aggregate to create the final arrangement. The first oscillating part of the

graph reflects the fluctuations of the two formed pairs before reaching the final assembly. The first interactions are again mainly hydrophobic.

The changes in the total number of inter-tropocollagens hydrogen bonds during the aggregation were monitored, highlighting the contribution of hydroxyproline amino acid, as well as the number of hydrogen bonds between protein and water. Also, the formation of salt-bridges was checked. The H-bonds plots for simulation with four tropocollagens, that are the more interesting, are reported in figure 36.

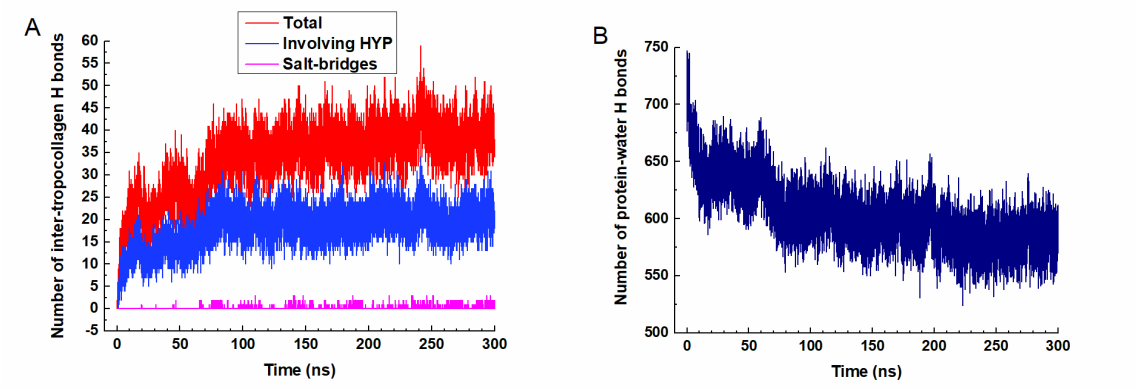


Figure 36. Hydrogen bonds formed during the simulation with four LC fragments. (A) Plot of inter-tropocollagens hydrogen bonds as a function of time; HB involving hydroxyprolines (light blue line) and salt bridges (magenta line) are also displayed. (B) Plot of protein-water hydrogen bonds versus time. Adapted from [65].

It is known that the absolute number of hydrogen bonds depends on the threshold fixed to reveal them; nevertheless, we achieved the same data trend also using other threshold values. As expected, the number of inter-tropocollagens H-bonds increases over time, as the collagen molecules get close. The fluctuations in the first part of the plot correspond to the dynamics of the pair coupling, already seen in the COM distance plot. It is worthwhile to discuss the contribution of hydroxyprolines, of which LC fragments are rich. In fact, a putative key role was assigned to this residue in collagen aggregation and fibril formation [86] [87]. The number of Hyp residues in our systems is between 14% and 16% of the total number of amino acids in T0, T1 and T2 fragments, whereas it is 7% in T3 (Table 2, in Material and methods). Observing the plot in figure 36A, the final contribution of Hyp residues is about 50% of the total number of hydrogen bonds, that is very high considering that the percentage of the total hydroxyprolines in the four tropocollagens is about 11%. In addition, we observe that Hyp residues make direct H-bonds with the glycine backbone through their side chains, rather than mediated by water molecules.

The number of salt bridges formed is very low, due to the poor number of charged residues and their arrangement along the sequence of the LC fragments.
Finally, the plot in figure 36B shows that the number of protein-water H-bonds decreases as the monomers are approaching, and the profile is specular to the one displayed in the plot of figure 36A.

5.1.1.2 Low ionic strength

Different results are obtained from simulations at low ionic strength. The decrease in the number of ions affects the collagen self-assembly: the aggregation of LC fragments is really difficult to occur. In both replicas, the complete assembly of three tropocollagens does not take place, but only partial association is obtained. As displayed in figure 37A, a shift of the chains along their axes takes place, in particular for T3, leading to an incorrect aggregation. As concerns the simulations with four tropocollagens, the formation of an aggregate occurred only in one replica (Fig. 37B), because, in the other one, only pairs are formed.

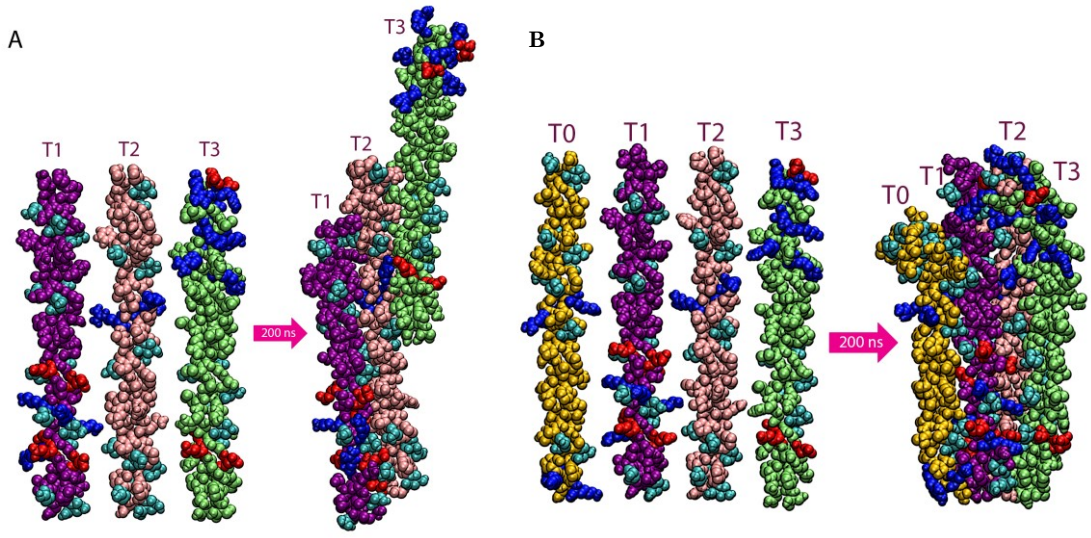


Figure 37. Aggregation of LC tropocollagens at low ionic strength. Van der Waals representation of the systems with three (A) and four (B) fragments before and after 200 ns of MD simulations. Triple helices are labelled from T0 to T3 and colored in different way. Hyp and charged residues are highlighted with residue type coloring (see Figure 31). Adapted from [65]

In the next picture, contact map of the last frame and interactions maps for the simulation with four tropocollagens are reported (Figure 38)

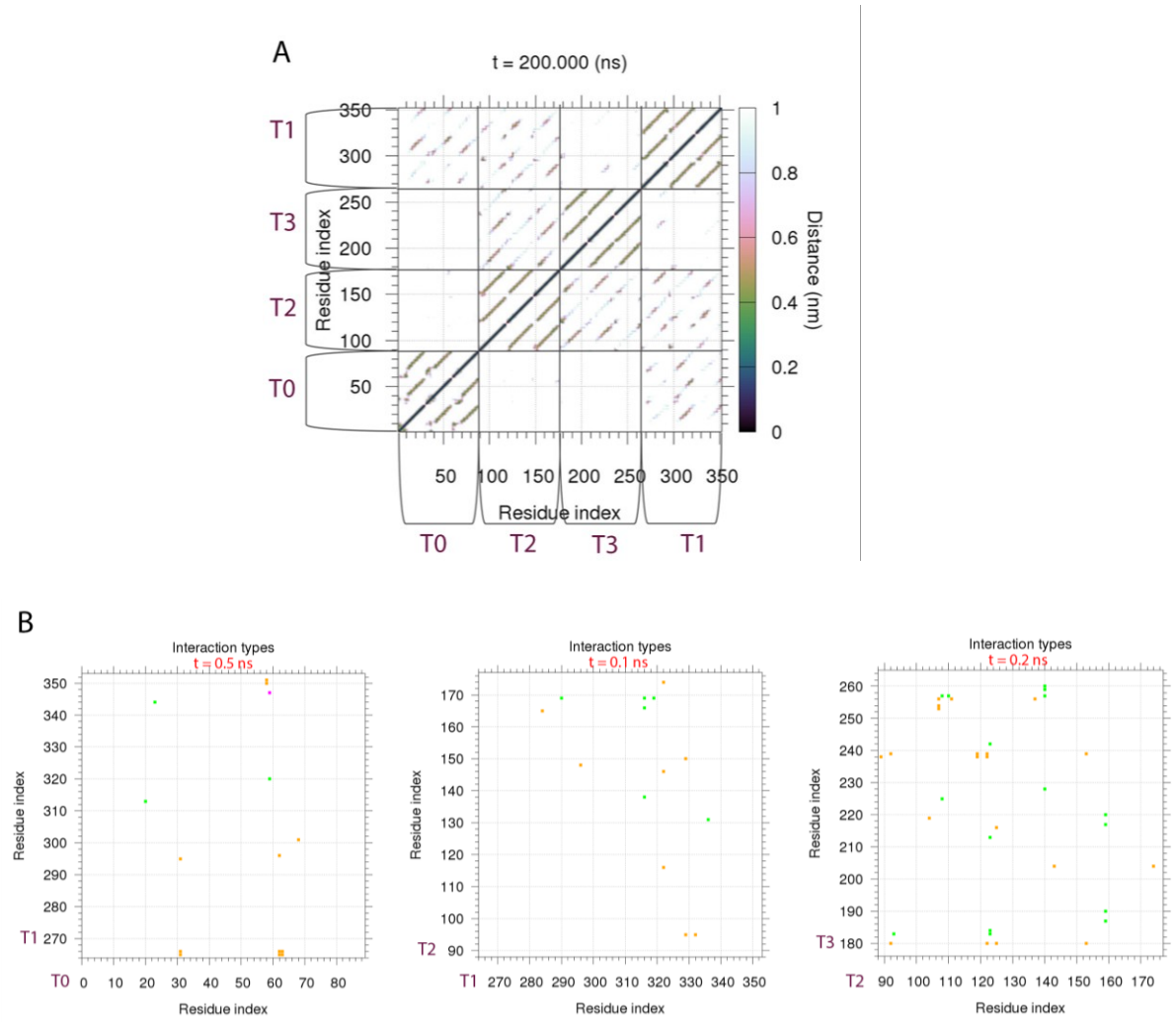


Figure 38. (A) Contact map of the last structure of the trajectory with four LC fragments. (B) Interaction maps of first contacts in T0-T1, T1-T2 and T2-T3 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

The fragments arrangement shown in Figure 37B and the complete absence of some intra-tropocollagen contacts in the contact map (Figure 38A) indicate the formation of a no more tubular but planar assembly. The interactions involve only the consecutive T0-T1, T1-T2 and T2-T3 pairs. The first interactions involve small portions of the fragments, that get close almost simultaneously, even if the complete assembly occurs sequentially, starting from T2-T3 pair, then T1-T2 pair and finally T0-T1 pair. The first contacts are mainly hydrophobic, but the polar contribution is also particularly relevant between T2 and T3, whose association occurs first and involve the whole fragments.

5.1.2 Aggregation of HC model fragments

In each simulation the aggregation of HC tropocollagens was tracked for 200 or 300 ns (according to the RMSD profile). Under physiological conditions, the assembly of two fragments occurred in all replicas, whereas the aggregation of three and four triple helices did not take place in 300 ns. It was noticed [88] that the lateral association of such short fragments is limited by their rotational freedom. For this reason, to make easier the association, the third and/or fourth tropocollagen was added to a previously assembled pair, placing each one at an initial distance of 2.0 nm. Simulations with two different pre-assembled pairs were tested to select the best initial structure, and the T2-T3 pair was chosen based on the observed aggregation in all replicas.

In the case of the simulations at low ionic strength, these aggregation problems did not occur, and the aggregation of separated fragments was observed. The first results, reported in the next paragraph, regard simulations at physiological salt concentration (0.1 M).

5.1.2.1 Physiological salt concentration

The first and the last structures obtained by the simulations at physiological salt concentration are shown in figure 39.

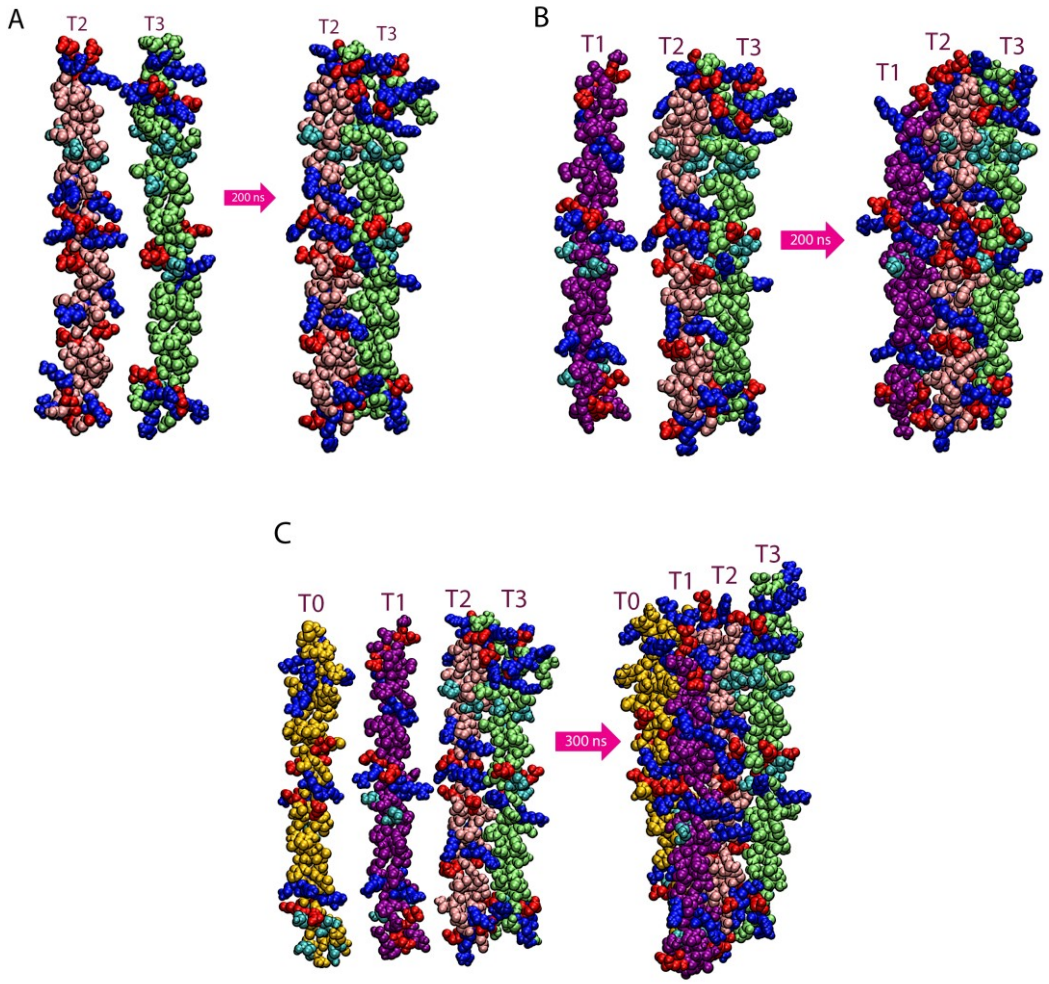


Figure 39. Aggregation of HC tropocollagens. Van der Waals representation of the systems with two (A), three (B) and four (C) fragments before and after 200 or 300 ns of MD simulations. Triple helices are labelled from T0 to T3 and colored in different way. Hyp and charged residues are highlighted with residue type coloring (see Figure 31). Adapted from [65]

While the aggregation of two and three tropocollagens is observed in all replicas, the aggregation of four tropocollagens is seen only in one of three replicas, since the assembly in the other two is only partial. The analysis of contact maps of the last frame, reported in figure 40, helps to recognize a planar shape in the three aggregated fragments and a more tubular one in the four assembled tropocollagens, as occurred in LC model.

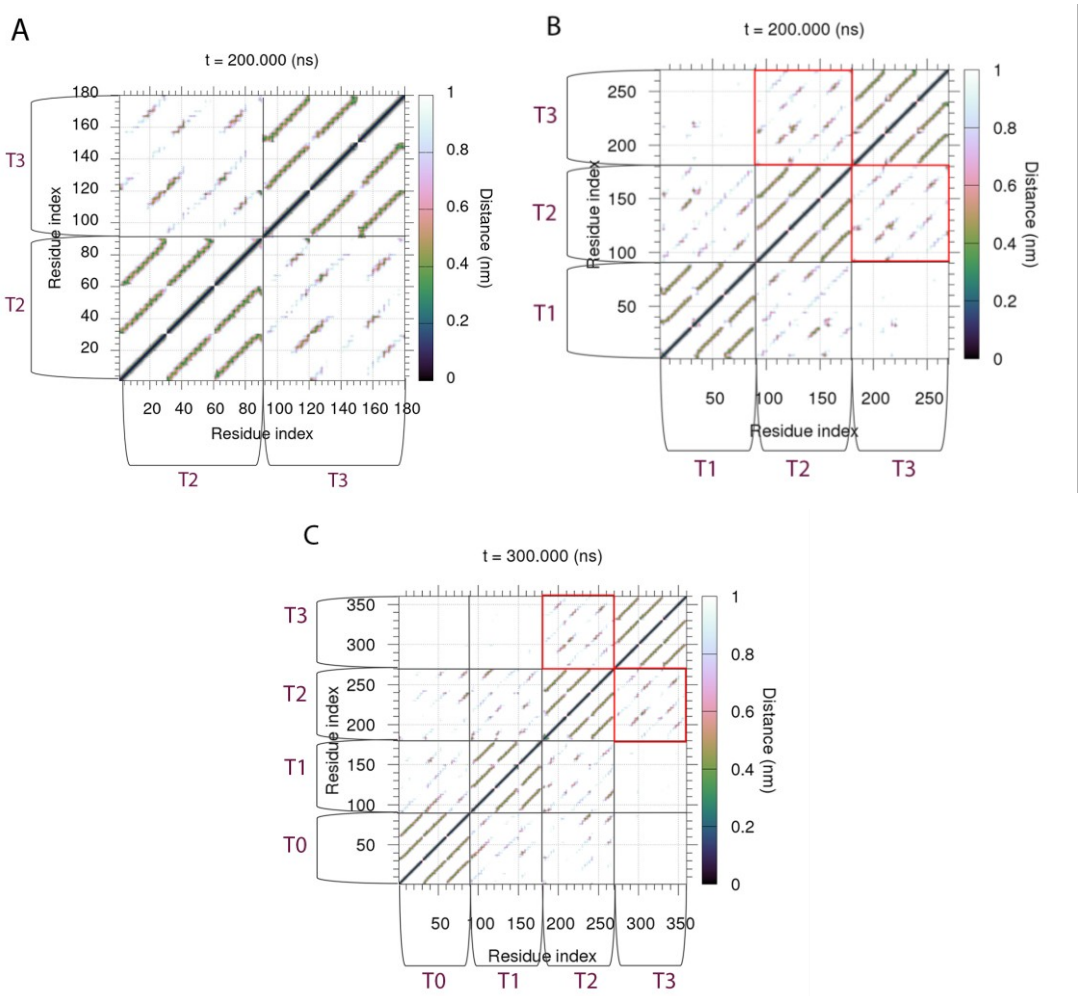


Figure 40. Contact maps of the last structure of the trajectories with two (A), three (B) and four (C) HC fragments. The red squares highlight the T2-T3 pair, already in contact in the starting structure. Adapted from [65]

COM distance plots and interaction maps of simulations with two and three fragments (figure 41 and 42), show a rapid aggregation. The aggregation of four tropocollagens, instead, is quite different, in fact, it takes place in a more sequential way (Figure 43). The inset of Fig. 43A, highlights that first T1 approaches T2, then T0 move to T1 forming the T0-T1 pair and finally the assembly is formed. The interaction maps confirm this sequential path: the first contacts are formed between T1 and T2 fragments (in 1.0 ns) and then between T0 and T1 tropocollagens (in 3.2 ns). In all the simulations, first contacts are more polar (purple and green spots) than in the lowly charged model, but the hydrophobic contribution is also relevant, especially in the T1-T2 pair, that possesses the highest percentage of hydrophobic amino acids and that are the tropocollagens that get close faster.

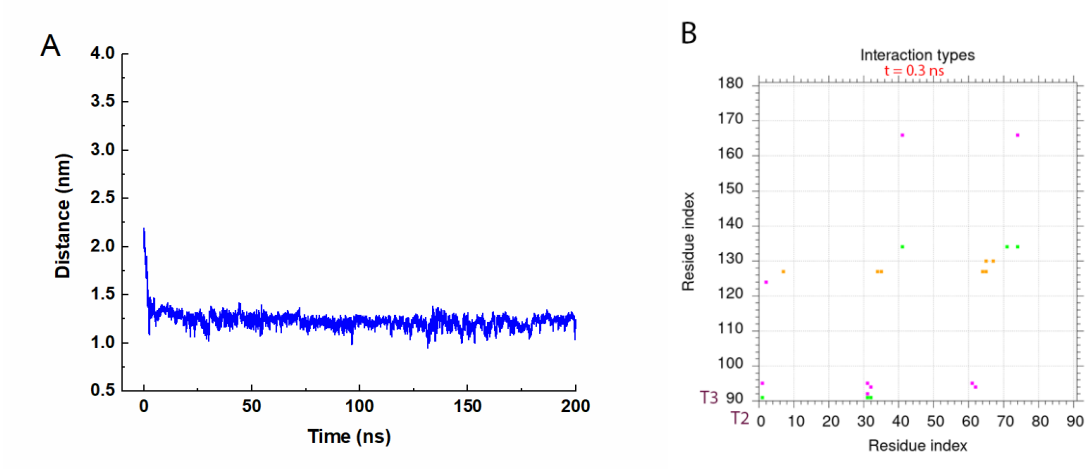


Figure 41. (A) Inter-tropocollagen COM distance during the two HC fragments (T2 and T3) trajectory. (B) Interactions map of first contacts of T2-T3 pair: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

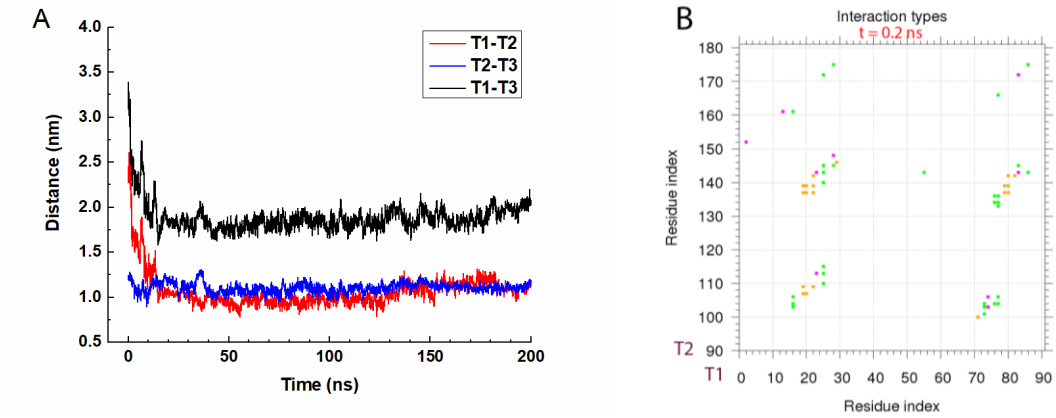


Figure 42. (A) Inter-tropocollagen COM distance during the three HC fragments (T1, T2 and T3) trajectory. (B) Interaction map of first contacts in T1-T2 pair: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

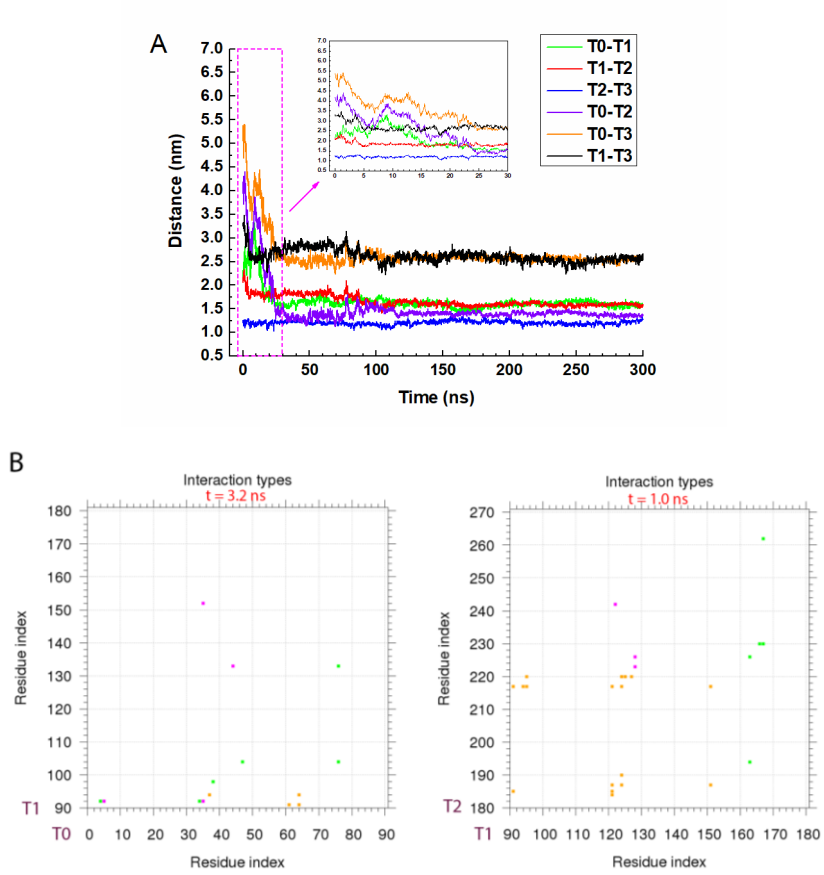


Figure 43. (A) Inter-tropocollagen COM distance during the four HC fragments trajectory. (B) Interaction maps of first contacts in T0-T1 and T1-T2 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

The number of inter-tropocollagen hydrogen bonds as a function of time, calculated from the simulation with four fragments, is reported in the plot of figure 44.

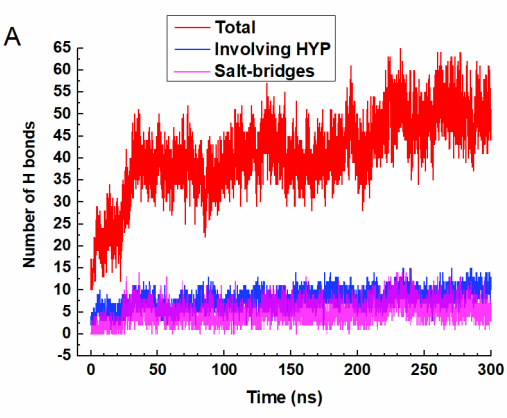


Figure 44. Inter-tropocollagen hydrogen bonds formed during the simulation with four HC fragments: HB involving hydroxyprolines (light blue line) and salt bridges (magenta line) are also shown. Adapted from [65]

The number of H-bonds is higher than the one reported for LC model, consistently with the higher number of polar residues that are able to make H-bonds. They are about 40% of the total residues for HC model, and about 35% for the LC one, whereas the solely charged residues in HC model are 23.9% of the total, and 9.0% in LC model. For the same reasons, also the number of salt bridges is much more considerable than in the LC model. As concerns the hydroxyproline contribution to the hydrogen bonds network, it has to be highlighted that the total number of Hyp residues in the HC model is low (6.4% of the total amino acids), nevertheless hydroxyprolines are responsible for the 20% of the total H-bonds. The contribution of Hyp residues to the H-bond formation is high also in the simulation with HC fragments, as previously observed for LC model.

5.1.2.2 Low ionic strength

The decrease in the number of ions shielding the charges of the system, greatly affects the HC fragments assembly, which seems to occur more easily than at physiological conditions. In fact, the association of three and/or four separated tropocollagens takes place without the requirement of a previously assembled pair, as shown in figure 45.

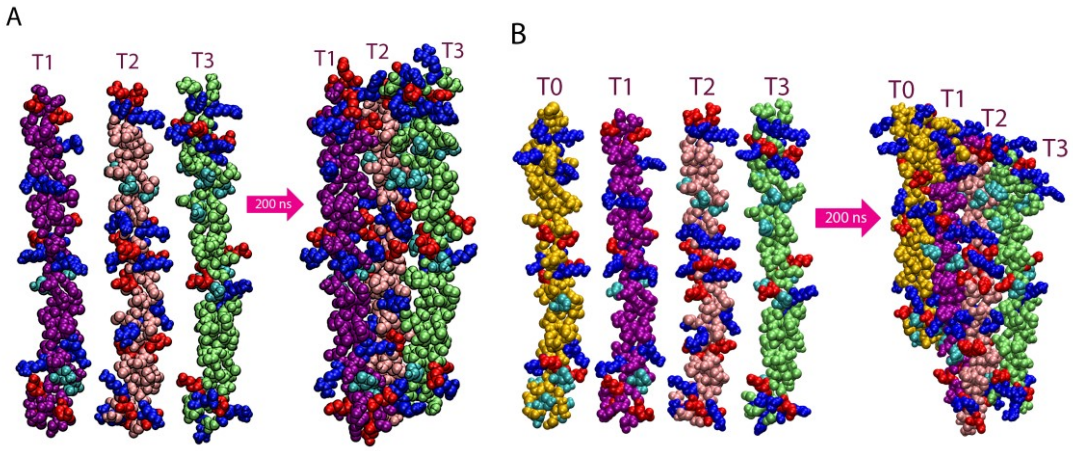


Figure 45. Aggregation of HC tropocollagens at low ionic strength. Van der Waals representation of the systems with three (A) and four (B) fragments before and after 200 ns of MD simulations. Triple helices are labelled from T0 to T3 and colored in different way. Hyp and charged residues are highlighted with residue type coloring (see Figure 31). Adapted from [65]

The three fragments aggregation occurs in both replicas, whereas the four monomers assembly takes place only one in two cases. The four tropocollagens aggregate with an incorrect staggering to allow the contact between oppositely charged residues (figure 45B). Contact maps (figure 46A and 47A) display the nearly planar arrangement of the assembly. In figure 47A, the shift of the chains can be noticed observing the absence of contacts in the lower right or upper left corners of T2-T3, T1-T2 and T0-T1 grid squares. The interaction maps of both simulations with three and four fragments (figure 46B and 47B) show that the first contacts are mainly polar, with a high presence of salt bridges.

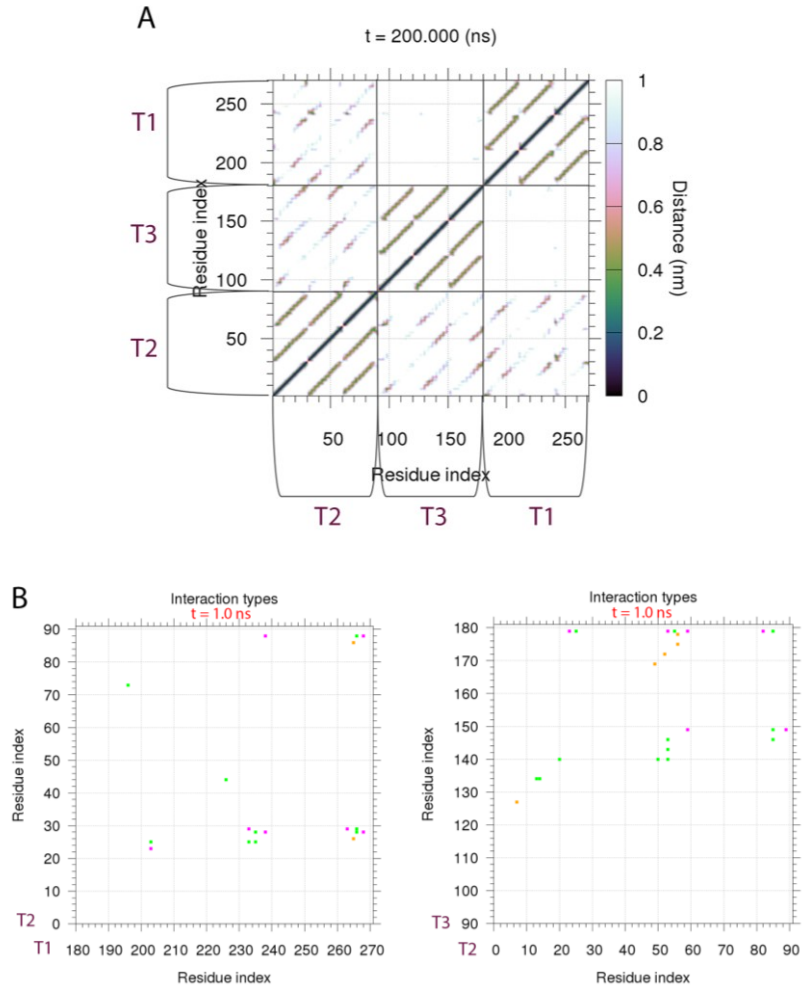


Figure 46. (A) Contact map of the last structure of the trajectory with three fragments. (B) Interactions maps of first contacts in T1-T2 and T2-T3 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

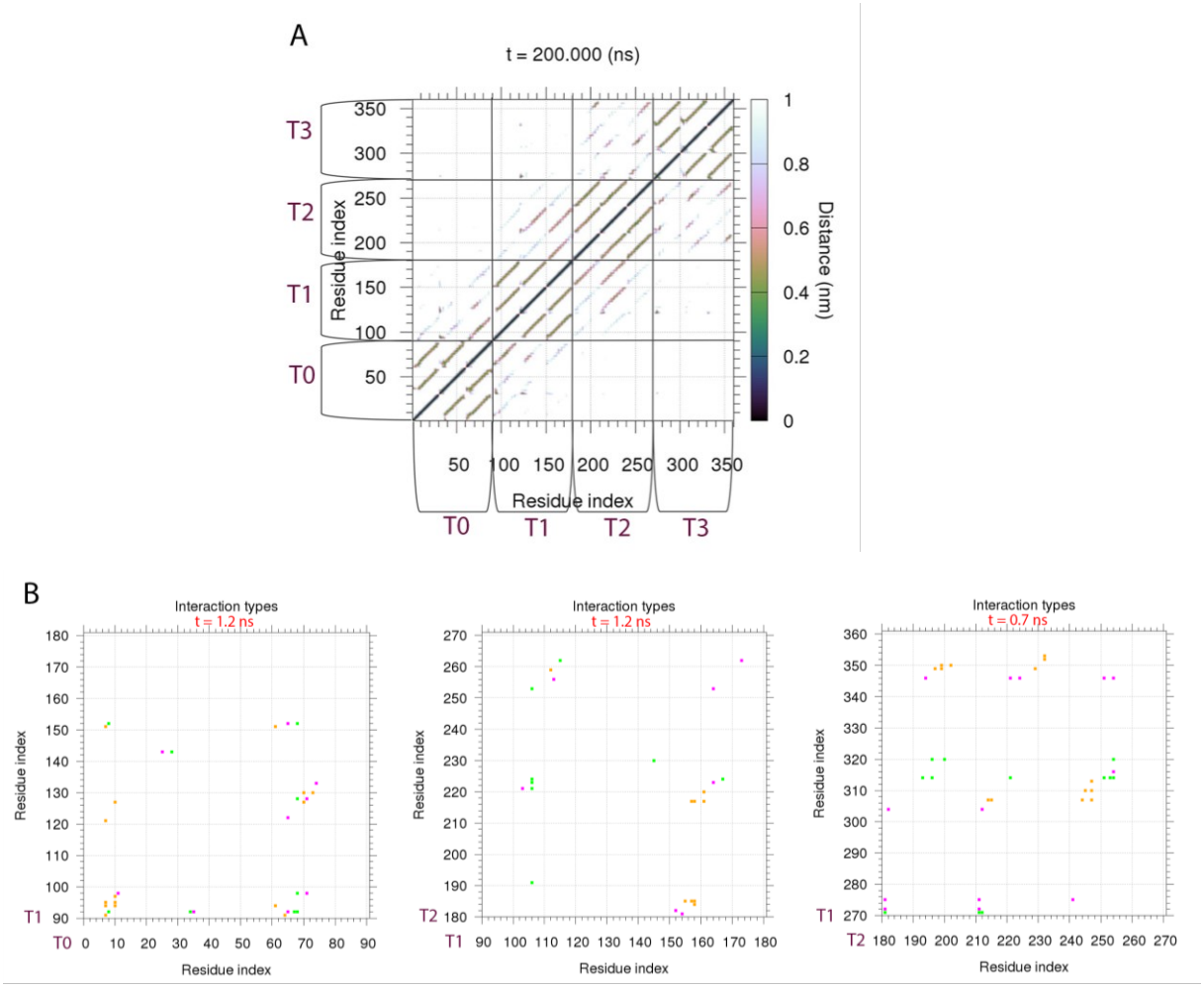


Figure 47. (A) Contact map of the last structure of the trajectory with four fragments. (B) Interactions maps of first contacts in T0-T1, T1-T2 and T2-T3 pairs: orange spots indicate hydrophobic interactions, green spots represent polar contribution and purple spots are salt bridges. Adapted from [65]

5.1.3 Differences in the aggregation between LC and HC models: insights into the assembly mechanism

The selection of two regions of the collagen sequence with different hydropathicity to study their aggregation allows to establish the influence in the assembly process of the aminoacidic composition. A previous research [86], in fact, hypothesized that the particular interactions between side chains and their size could influence the packing of triple helices. In addition, observing the whole crystal structure of rat tail collagen fibril at low resolution [13], we can realize that staggered fragments have a similar hydropathic profile, with the charged portions forming a sort of distinct blocks along the fibril. It was already noted [16] that fibrillogenesis resemble the protein folding process, for two reasons: the aminoacidic code regulates the three

dimensional structure and it is an energy driven process in which the entropy loss is stabilized by the enthalpic gain derived by favorable hydrophobic and hydrophilic interactions. This point of view suggests that the typical collagen staggering assures the best interaction between tropocollagens, approaching sequences with similar hydropathicity and, so, maximizing the number of positive contacts.

Concerning our simulations, it was shown that LC and HC models aggregate in different ways. The association of less charged regions occurs more easily at physiological salt concentration and it does not require a sequential order. At the same salt concentration, this aggregation path seems to be not possible for HC tropocollagens: a sequential assembly, in which the binding of two fragments promotes following bindings, is required and, so, it could be sequence-dependent. The first interactions between tropocollagens are mainly hydrophobic, suggesting an hydrophobic driven force at the base of assembly mechanism, according to Li and co-workers [89]. Concerning the aggregation of a whole fibril, we could suggest that the association of tropocollagens starts from less charged regions and then continue along the rest of the chains, hypothesizing a hydrophobic-driven aggregation. However, a sequential path could be needed to complete the whole assembly, allowing the correct match of highly charged regions, due to the relevance of electrostatic interactions at short distances. A decrease of the protein-water hydrogen bonds and a specular increase of the inter-tropocollagen H-bonds were observed as the fragments are approaching, confirming that self-assembly is a process driven by loss of water from the surface of monomers [10].

It was already claimed that hydroxyproline residue plays a key role in the stability of collagen, both inter- and intra-triple helix [5] [86]. Polarity of its hydroxyl group enhances the complementarity of the staggered collagen fragments [3]. Our study suggests that Hyp residue makes a high number of direct inter-tropocollagens H-bonds, involving, in particular, the carbonyl group of glycine, the most abundant residue in collagen. Water bridges mediated by Hyp appear only while the tropocollagens are getting close, as reported by literature [16], but are here hardly observed in the final aggregates.

For both models at physiological salt concentration shape of the assemblies with three tropocollagens is almost planar, whereas the addition of a fourth fragment induces a quite tubular rearrangement, consistent with the fibrillar structure [13].

Some differences in the aggregation mechanism are observed varying the salt concentration, and the two models act differently. At low ionic strength, the fragments associate only when they are able to match the charged residues of apposing tropocollagens. In fact, the loss of ions' shielding effect enhances the electrostatic interaction, that becomes the driving force in the self-assembly. The oppositely charged amino acids rapidly match, in this way leading to the association of tropocollagens, but resulting in an incorrect shifted aggregate.

From these premises, differences between the assembly mechanisms of LC and HC models are expected. The aggregation of LC fragments at low ionic force conditions is penalized, due to the poor number of charged amino acids and, in some cases, due to the lack of charges coupling between apposing tropocollagens. This lack can cause the shift of the fragments to put in contact oppositely charged residues along collagen chains. An example is observed in the simulation with three tropocollagens, where a wide shift of T3 takes place (Figure 37A). This shift, in fact, enables the favorable match between Asp/Glu25 (α 1/ α 2 chain), belonging to T3, and Arg13 in T2. According to this, it should be noted that the loss of ions penalizes in particular negatively charged amino acids, because only the chloride ions (not sodium ions) used to neutralize the system are present in solution, shielding partially the positive residues.

As regards HC model, the decrease in the ionic shielding promotes the association of HC fragments and sure enough single tropocollagens aggregate rapidly and simultaneously, not sequentially as observed at physiological conditions. At 0.1 M salt concentration, in fact, the ions slow down the charge coupling and the aggregation of HC tropocollagens can take place only at short distance, by way of displacement of opposite charged ions around the residues. Figure 48 shows an example of this mechanism found in the simulation with three HC fragments at physiological conditions: the displacement of a chloride ion away from arginine of T1 enables a salt-bridge formation between that arginine and a Glu residue of T3, allowing the aggregation of the two fragments.

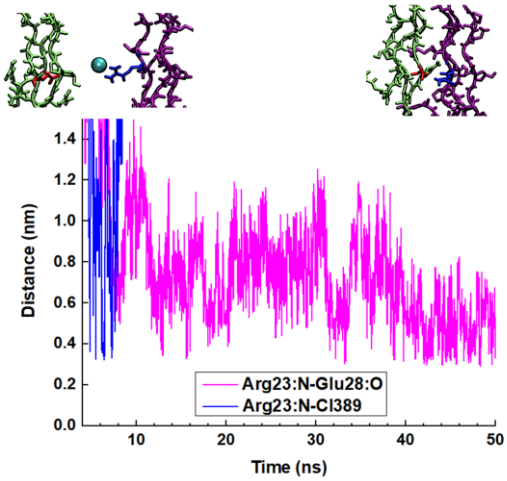


Figure 48. An ion displacement allows the salt-bridge formation between Arg23 of T1 and Glu28 of T3, during the simulation with three HC fragments at physiological salt concentration. Distance plot displays salt-bridge formation only after the chloride displacement away from arginine. The two images on the top show two representative frames indicating this behavior. Arg and Glu residues are highlighted by residue type coloring (see Figure 31). Chloride (cyan) is in Van der Waals representation. Adapted from [65].

In spite of the favorable conditions at low ionic strength, also the HC assembly is formed with a shift of fragments along their axes, forced by the coupling of oppositely charged amino acids. Observing the fibril crystal structure, it can be noted that the fibrillar staggering is strictly respected and it suggests that shifted assemblies obtained from simulations at low ionic strength are incorrect. In accordance with this hypothesis, experimental studies at low salt concentration have detected that collagen molecules form extremely thin microfibers, and the assembly process is fast [25]. It could be hypothesized that in low ionic strength conditions larger fibrillar structures are not formed due to the incorrect arrangements of tropocollagens.

It should be highlighted that *in vivo* collagen self-assembly takes place at physiological ionic concentration of about 0.1/0.15 M, where our results suggest an association that begins from lowly charged regions, is driven by hydrophobic interactions, and allows the correct packing.

5.2 UV measurements

Collagen self-assembly kinetics was monitored at three different pH values by means of UV-visible spectroscopy. The turbidity curves were recorded measuring UV

absorbance of the three solutions and then normalized to the maximum absorbance to obtain the aggregate fraction (AF) as a function of time (Figure 49).

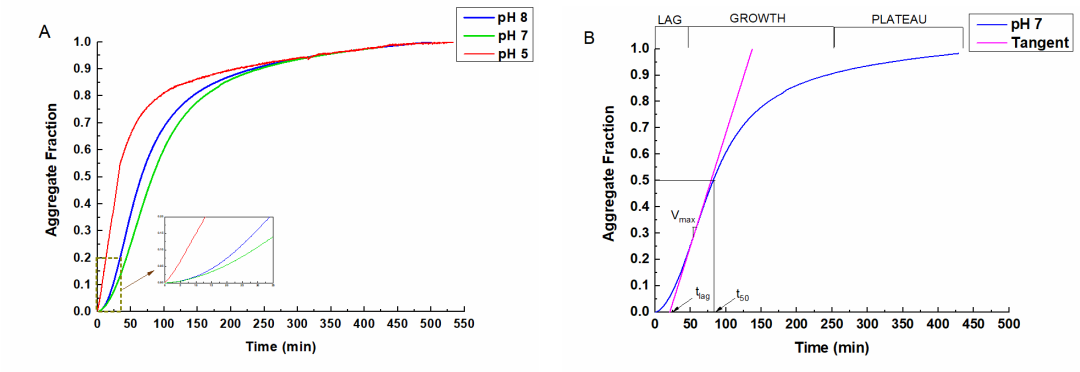


Figure 49. (A) Aggregate fraction as a function of time at three different pH values. The magnification of the plot between 0-35 min is shown in the inset, in order to distinguish the differences in the lag phase trends. (B) Aggregate fraction versus time at pH 7.0, in which three distinct phases are highlighted: lag, growth and plateau. Lag time is calculated by tracking the intercept of the tangent at the time in which the rate is maximum on the time axis. Adapted from [65].

All curves display a sigmoid profile, showing three distinct phases (highlighted in Fig. 49B for pH=7.0) characteristic of a filamentous protein aggregation in agreement with literature [85]. The first step is called "Lag phase" and it is characterized by no changes (or very little) in absorbance: this is the starting point of fibril nucleation. The second step, named "Growth phase", is identified by an increase in absorbance due to the growth of fibrillar cores. Finally, the last step is a "Plateau phase" characterized by the achievement of constant values of absorbance with time: the creation of three-dimensional networks of fibrils occurs. From these curves t_{50} and t_{lag} were calculated, as described in "Materials and methods" chapter, being related to the time in which half of fibrillar aggregates was formed and to the length of lag phase, respectively. The values of the two parameters are listed for each curve at different pH in the following table (Table 4).

pH	t_{lag} (min)	t_{50} (min)
pH 5	3.5±0.1	23.5±0.1
pH 7	21.0±0.1	82.3±0.1
pH 8	14.9±0.1	65.5±0.1

Table 4. t_{lag} and t_{50} by evaluated by turbidity curves. Adapted from [65]

The data suggest that the collagen self-assembly parameters critically depend on the pH conditions. The magnification of the plot representing the first 35 minutes of the process, as shown in the inset of figure 49A, allows to detect the differences in the lag phase trends. The collagen aggregation process at pH 7.0 exhibits the slowest growth phase and the longest lag phase, that can be correlated to a slower assembly with respect to the process developing at the other two pH values investigated. The self-assembly kinetics at pH 5.0 shows the fastest growth phase and the shortest lag phase: the half AF is formed in just 23 minutes, almost a quarter of the time to produce fibrils at pH 7.0. Finally, the assembly process at pH 8.0 displays shorter values than pH 7.0, revealing that the change of these parameters with pH is not linear, the maximum time values being achieved at neutral pH.

The differences measured in the aggregation kinetics, as previously described, could be induced by changes in the charges of the protein. Light scattering studies on type II collagen self-assembly showed that at extreme pH values, fibrils are not be able to form, due to the repulsive forces of the high net charge [25]. pH values slightly below or above neutrality may cause the protonation/deprotonation of some residues, reducing the surface charge of the protein.

In light of our MD simulation results, we could partially justify why the aggregation process is faster at pH 8.0 and pH 5.0 than at pH 7.0, following our hypothesis of an assembly process facilitated by less charged portions, and probably starting from them.

6. Results and discussion: study of collagen dehydration

In this section the results of the study of collagen dehydration, obtained from MD simulations and FTIR spectroscopy, combined with gravimetric measurements, are reported.

6.1 MD simulations

To simulate a dehydrated condition up to the first hydration shell, a 3 Å layer of water molecules was cut around the fully hydrated four-tropocollagens assemblies obtained by the previous study (see chapter "Materials and methods") (Figure 50).

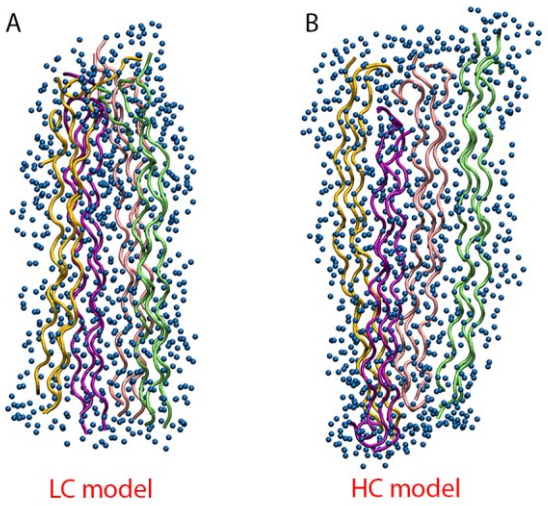


Figure 50. Representation of 3 Å water layer around LC (A) and HC (B) four-fragments aggregates at the beginning of simulations in dehydrated condition. Cartoon representation is selected for protein and cyan balls are water oxygens. Each tropocollagen is differently colored.

In this dehydration condition, we performed two MD simulation replicas (labelled S1 and S2) for both LC and HC models, checking their stabilization by RMSD plot (data not shown). The system configuration achieved from S1 simulation was then rehydrated, inserting it in a box of water. We preserved the water arrangement attained at the end of simulation in dehydrated conditions in order to monitor the

capability of water of the first shell to recover its original spreading. In this new condition, we performed two replicas of MD simulations (named S1.1 and S1.2) in order to study the reversibility/irreversibility of a dehydration-rehydration cycle on collagen microfibrils.

Figure 51 shows the final conformations of LC (A, B, C) and HC (D, E, F) assemblies, reached at the end of simulations in dehydrated (B) and rehydrated conditions (C), compared to the starting fully hydrated structures (A).

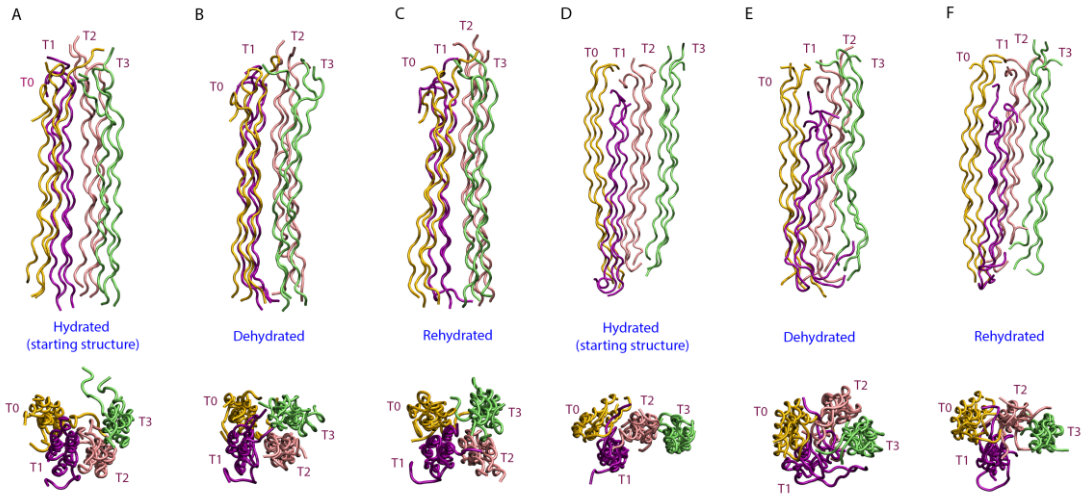


Figure 51. Cartoon representation of LC (A, B, C) and HC assemblies (D, E, F) along fibril axis (at the top) and in section view (at the bottom). Each tropocollagen is indexed from T0 to T3 and colored in different way. Starting structure (A, D), the last frame of simulation in dehydrated (S1) state (B, E) and in rehydrated (S1.1) one (C, E) are shown for both models.

As it can be observed, both LC and HC assemblies in dehydrated conditions become more compact. Section views at the bottom of figure 51 help to visualize this gain of compactness: T3 and T0 fragments of LC model appear closer than in the starting configuration, whereas the initial configuration of HC model goes from flatter arrangement to more tubular one in dehydrated state. In addition, in HC model, dehydration causes a partial unfolding in some portions of tropocollagens, bringing close single filaments of different triple helices, and, hence, resulting in a more compact final arrangement. The starting structures of both models seem to be recovered (even if not completely) in rehydrated state (Figure 51 C, F): the gained compactness is partially lost and the chains of HC model refold.

Contact maps, reported in figure 52 and 53, enable to visualize the interactions lost and re-established between pairs of fragments.

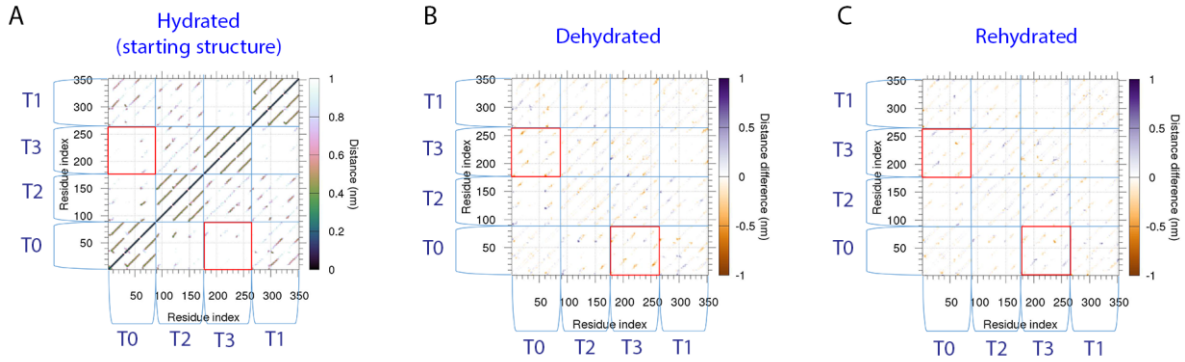


Figure 52. Contact map (A) and difference maps (B, C) of LC model. Contact map refers to hydrated starting structure, while difference maps refer to simulations in dehydrated (S1) and rehydrated (S1.1) conditions and represent the difference between the contact map of the last (at 500 ns) and the first (at 0 ns) structure. The red squares point out the variations in the distances of T0-T3 pair.

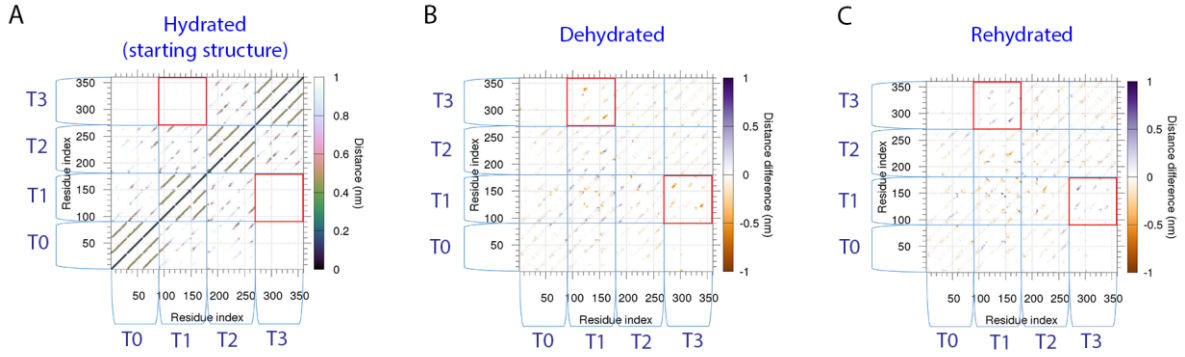


Figure 53. Contact map (A) and difference maps (B, C) of HC model. Contact map refers to hydrated starting structure, while difference maps refer to simulations in dehydrated (S1) and rehydrated (S1.1) conditions and represent the difference between the contact map of the last (at 500 ns) and the first (at 0 ns) structure. The red squares point out the variations in the distances of T1-T3 pair.

In particular, in figure 52 B, C and figure 53 B, C the difference maps, show the difference between the final and the starting contact map: orange spots indicate the approaching between two residues, and blue spots the moving away. As regards LC model, difference map in figure 52 B shows the appearance of interactions in the T0-T3 grid square and, in less extent, in T1-T3 one, highlighted by the presence of

orange spots indicating the approaching of tropocollagens. Conversely, figure 52 C shows a yellow-to-blue color inversion, indicating that the regions in contact in dehydrated state are turned back apart in rehydrated state. From these results, it is confirmed that the assembly gains compactness when it is dehydrated and loses it when it is rehydrated. The same is for HC model. In fact, in the difference map of Figure 53 B orange spots are visible, in particular between T1 and T3 and slightly between T0 and T3 tropocollagens, whereas difference map of figure 53 C shows a reversal of colors.

We can quantify this compactness, induced by dehydration conditions, calculating the radius of gyration of the backbone of collagen assembly around z axis (R_g^z), that is the axis of the fibril. Table 5 displays mean values of R_g^z, computed in the last 10 nanoseconds of all simulations.

R_g^z (Å)					
Simulation	**Hydrated**	**Dehydrated**		**Rehydrated**	
		S1	**S2**	**S1. 1**	**S1. 2**
LC	10.64 ± 0.12	9.56± 0.07	10.00 ± 0.08	9.65 ± 0.07	9.90 ± 0.06
HC	11.6 ± 0.11	10.27 ± 0.07	10.01 ± 0.06	10.72 ± 0.08	11.41 ± 0.12

Table 5. Means and standard deviations of R_g^z (Å) of fibril around z axis. These values are averaged on the last 10 nanoseconds.

In dehydrated systems, R_g^z decreases, in agreement with the previous analyses. In HC simulations the drop of R_g^z is greater than the one in LC simulations, due to the different starting structures, since LC assembly is much more tubular than HC one. In rehydrated systems of LC model, however, radius of gyration does not completely recover the starting value, suggesting that the gained compactness partially persists. R_g^z values of HC model in rehydrated state, instead, are closer to starting values, suggesting an almost complete recovery. It should be noted that we calculated the average of radius of gyration considering the whole length of the assembly, hence, R_g^z is strongly affected by unfolded portions of tropocollagens.

The study of protein-protein hydrogen bonds allows to evaluate the structure compactness discriminating the contribution of intra-tropocollagen and inter-tropocollagen changes (Figure 54).

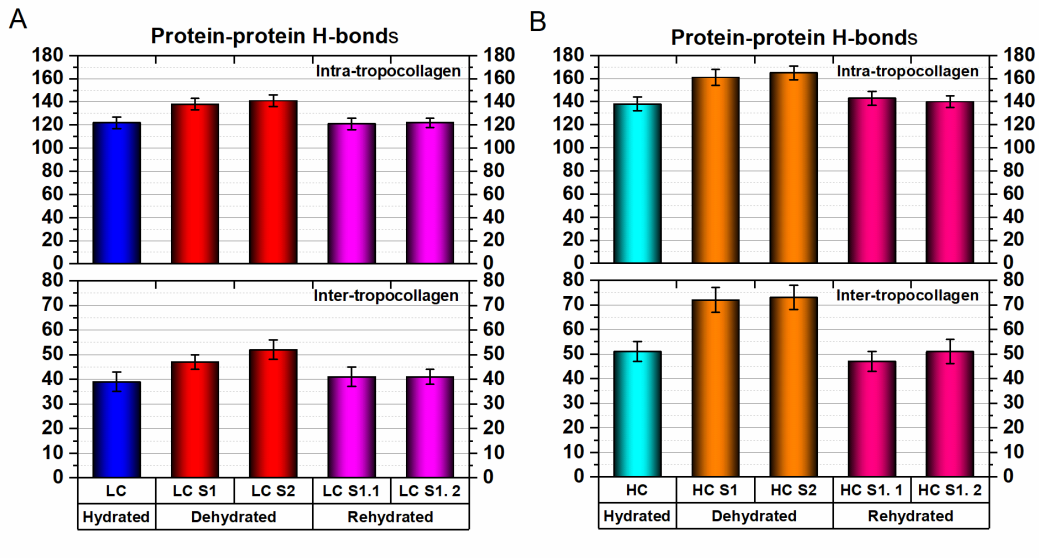


Figure 54. Mean number of inter-protein and intra-protein H-bonds, calculated from MD simulations of LC (A) and HC (B) models, during the last 10 ns. Error bars indicate the standard deviations.

As observed in figure 54, dehydration of collagen assemblies induces an increase of protein-protein hydrogen bonds (both intra- and inter-), reflecting the gain of a more compact configuration of collagen fragments. In particular, the number of inter-tropocollagen H-bonds rises by about 42% for HC model and 27% for LC model, the former possessing a higher number of charged residues that can interact. However, also the number of intra-tropocollagen H-bonds increases, by 18% and 14% for HC and LC models respectively, suggesting not only an approach of the collagen fragments, but also a greater compactness of the triple helix itself. The drop in H-bonds number, observed in rehydrated systems, indicate the loss of the gained compactness, suggesting a reversible process, in which the number of H bonds almost completely reverses back to the one obtained in fully hydrated conditions.

Despite of the greater compactness of dehydrated fibrils, several water molecules persist inside both HC and LC assembly, forming water bridges between the tropocollagens. This behavior can be observed calculating the volume maps of water occupancy, as reported in figure 55.

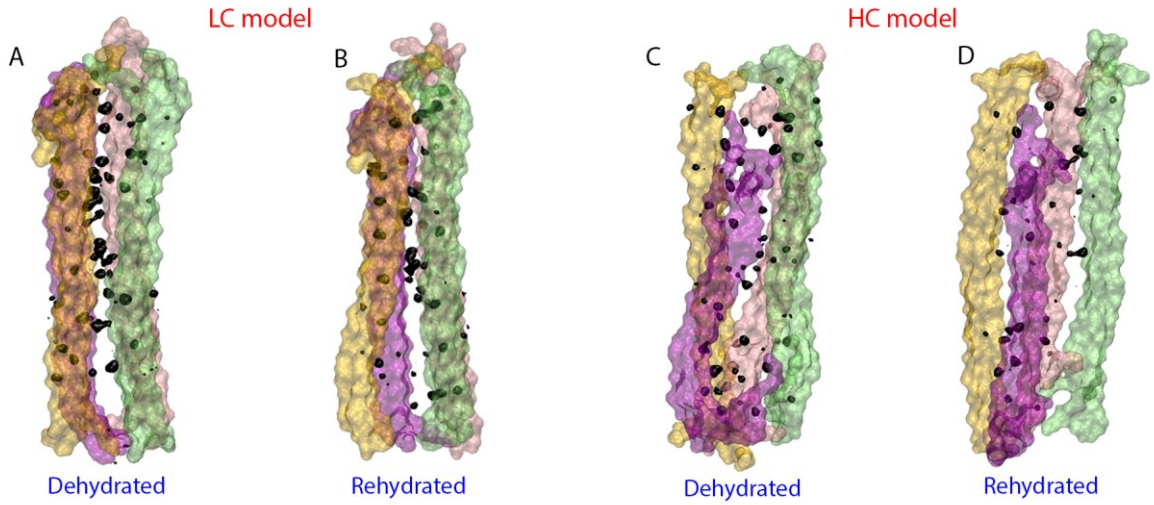


Figure 55. Volume maps of water occupancy with a threshold of 60% (black spots) computed and averaged on the last 10 ns of S1 and S1.1 simulations. Transparent surface representation is used to illustrate the last structure (only backbone) of trajectories of LC (A, B) and HC (C, D) models. Each tropocollagen is differently colored.

In dehydrated conditions, calculation of the volume maps of water occupancy with a quite high threshold of 60% indicates a high persistence of water that fills the inner cavities of both LC and HC assemblies, displayed as several black spots located inside the protein structures. After rehydration, volume maps of water occupancy in LC model still display trapped water inside the assembly, even if in less quantity due to the enlargement of the collagen fragments at the C-terminus. On the contrary, the rehydration of HC assembly, which induces the loss of the compact tubular shape, makes the trapped water molecules to let out. It is noteworthy that we computed occupancy volume maps for all the water inside the box, but very few external waters are shown when a quite high occupancy threshold is selected, in agreement with the cited ultraslow motions and the reported high values of residence time for water inside protein cavities [90][91][92].

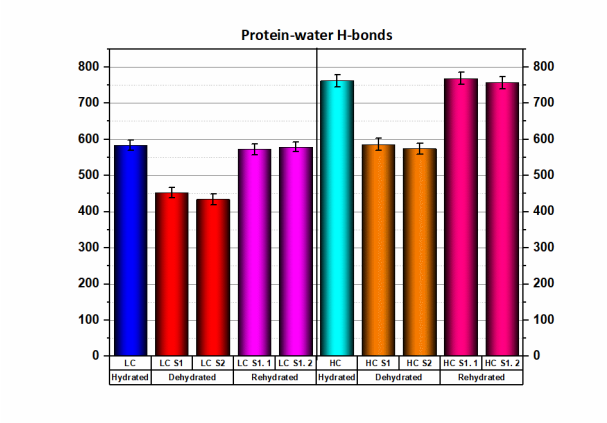


Figure 56. Mean number of protein-water H-bonds, calculated from MD simulations of LC and HC models, during the 10 last ns. Error bars indicate the standard deviations.

The mean number of protein-water H-bonds (figure 56) displays a reverse trend compared to the protein-protein one shown in figure 54. The analysis of occupancy volume maps suggests that the decrease of protein-water H-bonds in dehydration conditions is not caused by the loss of the inner water, but is due to the loss of outer water. Rehydration of both models induces a recovery of these H-bonds.

These results suggest that the first water shell that totally covered the protein at the beginning of simulations, in dehydrated conditions undergoes a rearrangement. In order to investigate on this behavior, the mean number of water molecules lying in a neighborhood of 3 Å around the protein was computed (Table 6)

Number of water molecules in a 3 Å layer					
Simulation	Hydrated	Dehydrated		Rehydrated	
		S1	S2	S1. 1	S1. 2
LC	760 ± 16	433 ± 9	411 ± 10	714 ± 15	723 ± 14
HC	927 ± 20	544± 12	530 ± 13	924 ± 20	935 ± 20

Table 6. Mean numbers and standard deviations of water molecules belonging to a 3 Å layer around the collagen assembly. These values are averaged on the last 10 nanoseconds.

As observed in table 6, the decrease in the number of water molecules in a 3 Å layer indicates that dehydration induces a rearrangement of water shell around protein, that is no more homogeneously distributed. In fact, as can be observed in figure 57,

water molecules form clusters, mainly covering charged residues and preferring to interact with itself than with the protein.

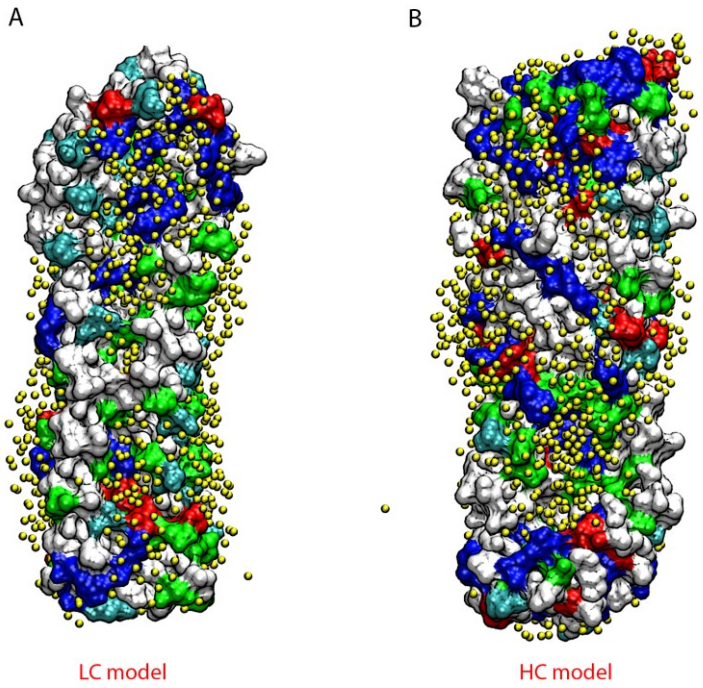


Figure 57. Surface representation of the last LC (A) and HC (B) structures of the dehydrated simulations. Yellow balls represent oxygens of the water. Both surfaces are colored by type of residue: polar amino acids are green, Hyp residues are cyan, basic amino acids are blue and acidic residues are red. Water is clustered mainly around charged zones.

In agreement with this behavior, previous simulations on myoglobin at 0.35 h revealed that the hydration shell is not a homogeneous layer due to the formation of water clusters around charged regions of protein, leaving 37% of protein surface uncovered [30]. We can explain our results considering two effects. First, the formation of ordered clathrates around hydrophobic regions on protein surface with such few water molecules should be too entropically unfavorable compared to the possibility to make H-bonds in flexible clusters. Second, water arrangement could be affected by an electric field caused by ionizable groups acting on dipolar water molecules and by dipole molecules themselves [93]. Charged regions are favored also in our simulations, according to the reported ranking of protein-water H-bonding potential which indicates that NH_3^+ and COO^- groups have the highest value among the possible H-bonds acceptors/donors of protein sidechains [94].

The addition of water molecules in rehydrated state enables the recovery of the monolayer that uniformly covers the protein.

In light of this water rearrangement, we tried to evaluate water mobility, computing the survival probability of a 2 Å layer of water molecules (i.e. the closest water to the protein) in the last 10 nanoseconds of trajectory and calculating the relaxation times. Decay curves of SP(t) are displayed in figure 58 and relaxation times τ, determined from exponential fit (see the chapter "Materials and Methods"), are reported in Table 7.

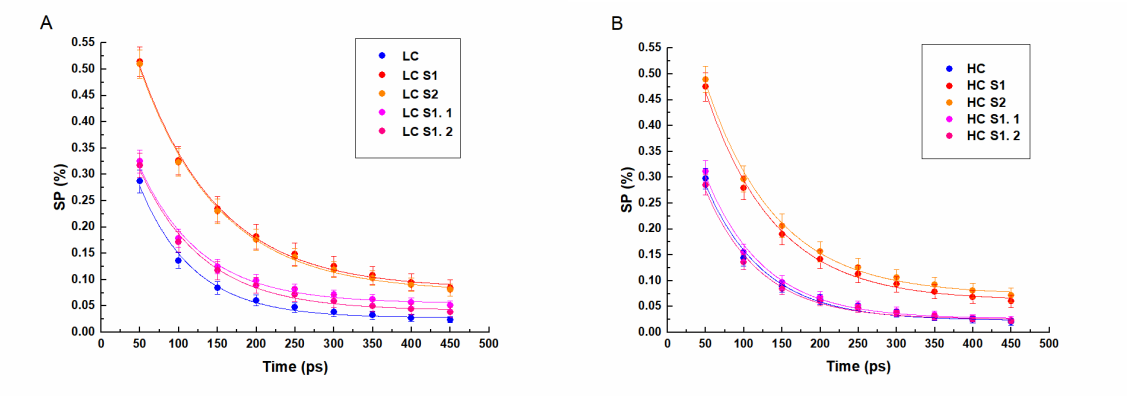


Figure 58. Survival probability as a function of time of a 2 Å shell of water, calculated in the last 10 ns of each simulation of LC (A) and HC (B) models. The circle symbols indicate the mean value of SP(t), error bars represent the standard deviations and the lines display the exponential fits.

Relaxation time τ (ps)					
Simulation	Hydrated	Dehydrated		Rehydrated	
		S1	S2	S1. 1	S1. 2
LC	69.6 ± 5.5	100.9 ± 5.9	102.4 ± 5.3	79.7 ± 6.7	81.9 ± 6.7
HC	75.6 ± 6.2	90.5 ± 5.2	92.1 ± 4.9	75.8 ± 5.2	71.8 ± 5.5

Table 7. Relaxation times and standard deviations, computed from exponential fits of the SP(t).

In both models, the values of survival probability are higher and show a slower decay in dehydrated state than in initial hydrated state, causing an increase of water relaxation time. We can explain these results considering two factors: the

slower motions of the inner water (included in the calculation) and a more reduced interchange of the first-layer molecules with the external layers [93], due to a limited number of water molecules around protein. However, the first factor is not crucial because the inner water is present also in rehydrated state and in this condition relaxation times and decay profiles of SP(t) are close to the ones in the fully hydrated state. Therefore, water trapped inside the tubular assembly does not affect very much the evaluation of relaxation time. The increase of relaxation times in dehydrated state can be explained also considering that the water clustering around charged protein groups tends to immobilize water molecules, so slowing down their motions. In addition, we can note that the differences in relaxation times between dehydrated and initial hydrated state are more marked in LC model than in HC one. This could be caused by a higher presence of hydrophobic regions close to the H-bonding water sites on protein surface, which has been shown to reduce the possibility for the water to move away from those sites [95].

In light of all the results we can suggest that rehydration process enable collagen fibrils to recover their starting structures. In particular, HC assembly recovers completely the initial flat shape. In LC model, instead, the gained tubular shape partially persists after rehydration, but this conformation is not very different from starting structures. Also, water recovers its initial mobility and distribution. The moderate hydration level of 0.42/0.49 h, which corresponds to our dehydrated state, seems to induce only reversible changes in the protein conformation.

6.2 Experimental measurements: gravimetric and FTIR studies

We studied the structural changes in lyophilized rat tail collagen sample induced by water removal and restitution along a dehydration-rehydration cycle. The use of saturated salt solutions which are able to control the known water activities a_w enabled to attain the desired water content in the protein sample to follow the desorption/adsorption process. According to the sorption method, the amount of water vapor which is sorbed by the sample determines different h values along the

adsorption and desorption isotherms. The sorbed vapor amount is measured by the mass increase of the sorbent (the protein) with the aid of a very sensitive balance. In figure 59, the plot of the hydration level, achieved by the protein sample, as a function of environmental water activity, provided by the salt solutions in the dry box, is reported:

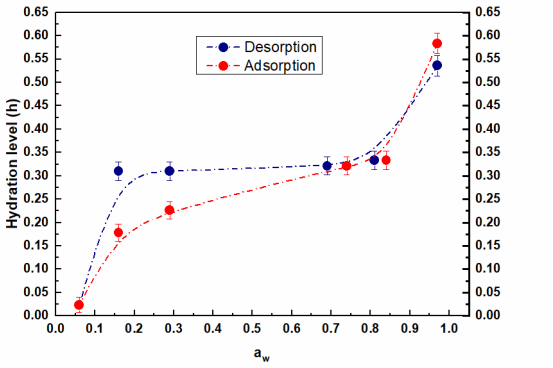


Figure 59. Desorption and adsorption isotherms obtained by means of gravimetric measurements. The dashed lines are a guide for eyes. Error bars represent the propagation of balance error.

Hydration value, h, expresses the ratio of mg water/mg dry protein and is estimated as explained in "Material and Methods" chapter. Error bars represent the propagation of balance error (± 0.0001). Different hydration degrees were obtained by means of the employed hydrating solutions, in particular, h was made to vary in the range 0.02–0.6. Moreover, the lowest value, 0.02, is well below the threshold value chosen for MD simulations i.e. 0.42/0.49. In addition, 0.53 ± 0.02 h is the starting hydration value, corresponding to the maximum water activity (0.97) in correspondence of which the protein is totally covered by water molecules. This value is not so far from the ones of simulations (0.42/0.49 h), corresponding in both LC and HC models to a water layer of 3 Å thick around collagen fragments. At the end of desorption/adsorption cycle, the collagen molecule adsorbs slightly more water molecules than at the starting of the process. The desorption curve shows the typical shape of the type I isotherm characteristic of microporous adsorbent, in which we can distinguish the characteristic knee in the a_w range $0.16 - 0.06$, while the subsequent adsorption curve displays the behavior of the type II isotherm. Type II isotherm indicates that the adsorbent is not porous and in the case of collagen such a change in the isotherm shape could indicate a packing of the fibers as a

consequence of the dehydration treatment. Both the curves display a similar trend for high a_w values up to $a_w = 0.7$, followed by different evolution up to $a_w = 0.06$, responsible of the hysteresis behavior. Hysteresis is commonly observed in the sorption isotherms of proteins and some explanations have been given in literature [19]. One of these hypotheses is concerning the changes attained by protein conformation when it is brought to a low hydration level. This is slowly restored with respect to the rate of water adsorption. Another explanation concerns the limited number of sites for water condensation on the protein surface: water needs charged sites acting as nuclei for interaction with protein surface. The plateau region of the plot, extending in the a_w range $0.84 - 016$, and corresponding to the larger hysteresis effect, can be due to the adsorption of water molecules during the rehydration process occurring only by the addition of water molecules to those clusters already formed around charged sites [19]. This last hypothesis could be in agreement with our results: the rearrangement of water molecules around charged zones of collagen was already seen in MD simulations at intermediate h values (0.42/0.49). During the desorption phase, the more external water molecules belonging to these clusters are removed and, at the end, very few molecules are surviving, interacting mainly with charged protein groups. At the start of the adsorption phase, therefore, charged sites are occupied by the remaining water molecules or buried and therefore not accessible, due to conformational rearrangements of tropocollagens in the fibrils induced by the low hydration. As a consequence, during the subsequent adsorption process not all water is restored at the starting amount: the water molecules could only join to the former clusters. In our collagen, we observe this behavior up to $a_w = 0.7$, but above this threshold the adsorption isotherm follows the same trend of the desorption curve and the hysteresis is reduced. We hypothesize that internal water cannot be totally stored, for the reasons just explained, and larger amount of external water compensates for this loss.

An analogous analysis was performed by means FITR spectroscopy with the aim to obtain water sorption isotherms from the study of OH stretching band of water interacting with protein. It is well known in fact that the area of such a band (3600-3000 cm-1) is proportional to the water sample content. v(OH) areas were calculated starting from FTIR spectra of collagen film shown in figure 60.

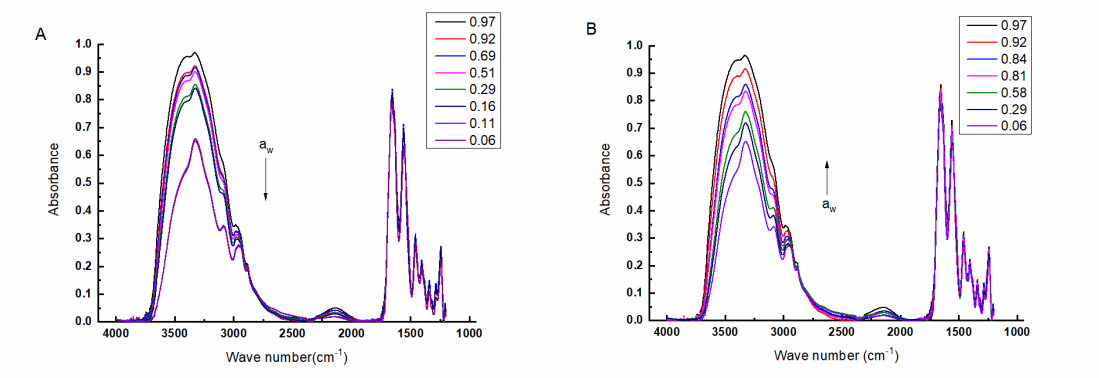


Figure 60. FTIR spectra of collagen film, recorded at different a_w values during dehydration treatment from a_w= 0.97 to a_w=0.06 (A) and rehydration treatment from a_w= 0.06 to a_w=0.97 (B). The colors of the spectra lines are corresponding to different a_w values as displayed in the caption.

From the curves of figure 60 we obtained water spectra of v(OH) region, by subtracting from each spectrum the spectrum of the dry sample, as described in "Materials and methods" chapter. In Figure 61 only the region in the range of v(OH) feature 4000-2500 cm^{-1} is shown. (Figure 61).

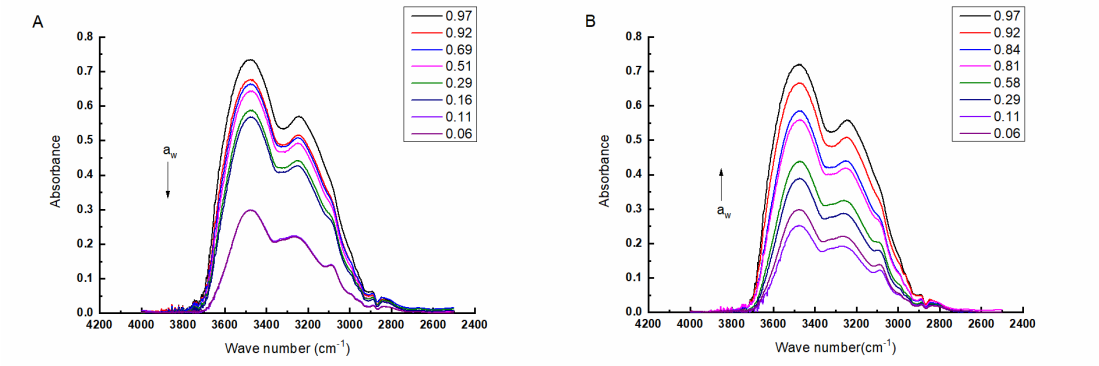


Figure 61. FTIR spectra of water stretching bands (4000-2500 cm^{-1}), recorded for the sample prepared at different a_w. In the left panel (A) is shown the evolution of the spectrum during the dehydration treatment, in the right one (B), during the rehydration process.

We calculated the v(OH) areas of the curves of figure 61 and thereafter we plotted the OH areas as a function of water activity (figure 62).

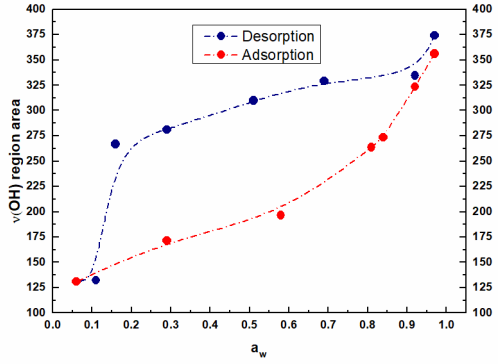
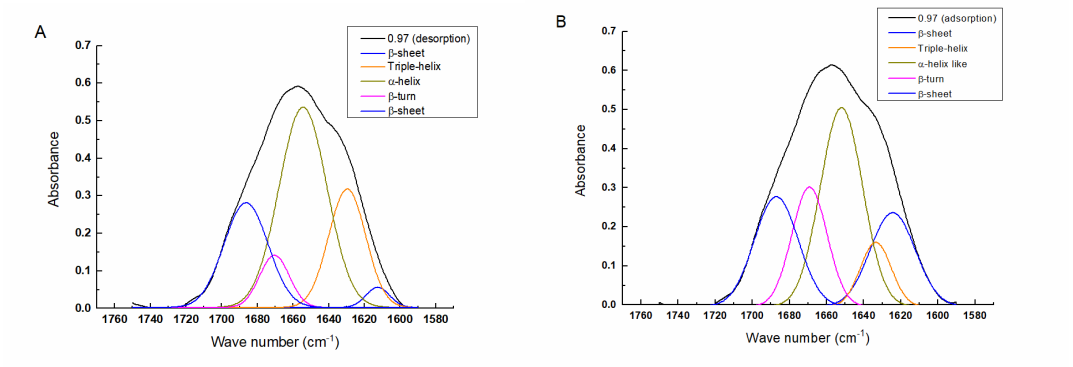


Figure 62. Desorption and adsorption isotherms deduced by the areas of v(OH) bands. The dashed lines are a guide for eyes.

Desorption and adsorption isotherms shown in fig. 62 exhibit a large hysteresis effect extending on the whole activity range spanned whose causes are the same discussed above. In this case the hysteresis loop regards not only the plateau region but the whole water activity range. The desorption curve displays the typical shape of type II isotherm, resembling quite well the gravimetric curves of figure 59, whereas the adsorption curve shows a different profile, closer to type III isotherm, in which a knee at low water activity is absent. The presence of the knee in adsorption curves, as in type II isotherms, is indicative for the formation of water clusters starting from single isolated water molecules on charged sites by adding water molecules to other ones linked to protein, [19]. In this case, instead, the step corresponding to the beginning of cluster formation is absent, so this could suggest that water molecules could suggest the propensity of water molecules to fill uncovered regions along the collagen molecules, rather than to form a layer at a time. In addition, curves like type III or V isotherms indicate weak interactions between adsorbent and adsorbate. These results could suggest that most of charged sites are buried and, being water molecules not able to percolate to build clusters, only isolated water molecules can fill uncovered and not buried polar sites. In this second experiment, the behavior of these water molecules is monitored. For these reasons we observe a hysteresis loop more broadened in figure 62 with respect to that shown in figure 59. The cause of this phenomenon could be a rearrangement of collagen molecules, like the compaction of fibrils induced by dehydration process, as observed in MD simulations. At the end of the adsorption process we can observe that the isotherm reaches a comparable area value of that measured at the start of the desorption process (in the limit of the experimental error). As observed in the gravimetric isotherms (figure

59), it could be possible that larger amount of external water compensates for the loss of internal water. It should be remembered that it is not possible to deduce the hydration values from FTIR area peaks and/or to simply relate them to those obtained by gravimetric measurements. Difference in the profiles of isotherms obtained by means of the two techniques can be attributed to the difference in the physical state of the samples in the two experiments at the end of which, however, the dehydration process leads the collagen to reach comparable hydration values, indicating a general agreement in the results. In addition, the very low intensity of the ν(OH) band and its small area at $a_w = 0.06$ suggest a very low water content, and therefore a strong dehydration, like the one recorded with gravimetric measurements. The hysteresis loop of the sorption curves achieved by FTIR measurements and gravimetric study could derive from different packing of collagen molecules in the two samples submitted to the dehydration process.

From FTIR spectra, in addition, it was possible to determine quantitative and qualitative information about the conformation of the collagen, through the analysis of Amide I band. The study of Amide I band was performed on FTIR spectra recorded at $a_w = 0.97$, before and after dehydration at $a_w = 0.06$ (figure 63).



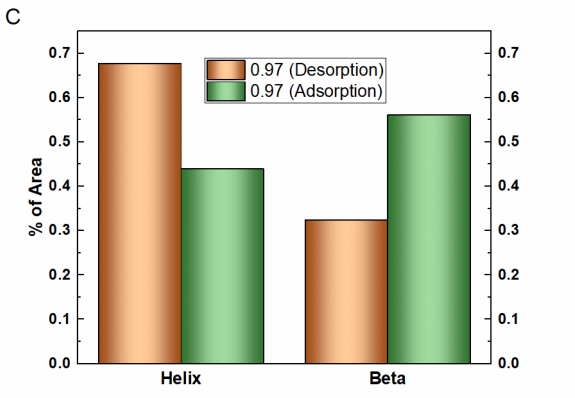


Figure 63. Amide I band deconvolution in Gaussian components, extracted from FTIR spectra at $a_w = 0.97$, before (A) and after (B) dehydration at $a_w = 0.06$. Each gaussian component related to a different secondary structure is differently colored. (C) The percentage amount of the beta and helix conformations were obtained from the corresponding Gaussian band area as related to the total band area.

The Amide I band deconvolution in Gaussian components allows to investigate the secondary structure composition, recognizing typical sub-bands whose areas and peak positions are related to the amount and different type of secondary structures, respectively. As shown in the figure 63, the sub-bands can be divided into two main categories corresponding to helix and beta sheet/turn conformation. Two helix components were identified in Amide deconvolution, namely, alpha and triple helix. In collagen, alpha helices structures are absent, however triple helix shows helical parameters very similar, hence, the alpha helix component can be called alpha helix-like structure and, as well as the triple helix band, it characterizes the triple helix structure, typical of tropocollagen. Secondary structures like beta-sheet are also absent in collagen molecules, therefore the presence of beta sheet and beta turn components in collagen Amide I band, indicates the presence of H-bond interactions between close fragments. These interactions provide hints on the presence of aggregates and, so, in this case, are indicative for collagen fibrils. The analysis of the two plots of figure 63 A and B, reveals an increase in the amplitude of beta turn (at 1670 cm^{-1}) and of beta sheet bands at the lowest frequency (at 1620 cm^{-1}) and a decrease in the amplitude of helix component bands, as a consequence of dehydration treatment at $a_w = 0.06$. In order to better compare the changes induced by dehydration process, the percentage areas of helix bands and of beta bands were calculated (Figure 63 C). The histogram clearly shows that the percentage area of beta structures overtakes the one of helix components at $a_w = 0.97$ after dehydration process, differently from the starting conditions, in which the area of helix components was higher than the ones of beta. The increase of beta structures with respect to the helix ones can be explained by considering an increment of fibril compactness or a growth of other fibrillar aggregates. This result could contribute to enhance the explanation about the hysteresis loop shown in figure 59 and 62. The inner water molecules hydrating the sample before dehydration process could be not totally recovered after dehydration-rehydration treatment due to an increase of collagen compactness induced by dehydration conditions. This hypothesis is in

agreement with results of MD simulation, where we have seen that deprivation of water molecules from protein surface induced a more compact structure in collagen fibrils. These changes in the arrangement of collagen molecules detected by MD simulations are reversible, whereas FTIR experiments show irreversible conformational changes, as demonstrated by the differences in Amide I sub-bands composition at the same a_w before and after dehydration. However, it has to be highlighted that in MD simulations we induce in the system a soft dehydration, corresponding to the removal of water up to the first shell, whereas in FTIR and gravimetric measurements collagen sample is submitted to a very strong dehydration. In fact, it is noteworthy that the dehydration of sample from liquid phase to dry state, in which, however, the sample is still covered by a shell of water, does not cause irreversible changes in collagen molecules. A strong dehydration, instead, removing almost all water molecules (except the structural ones that are strongly bound to the protein), induces irreversible changes in collagen fibers.

7. Conclusions

In this PhD work, two main studies concerning the collagen molecule have been performed by means of different techniques. First, collagen self-assembly mechanisms have been studied through molecular dynamics simulations and UV measurements; second, effects of collagen dehydration have been investigated through MD simulations and FTIR spectroscopy.

By means of molecular dynamics simulations the influence of different hydropathic profiles in the sequence on collagen aggregation was investigated. The long chains of a tropocollagen molecules, in fact, show polar and hydrophobic residues allocated with a certain periodicity and staggered fragments have similar hydropathic features [10]. For this reason and due to the length of tropocollagen sequence, too high to perform an atomistic MD study, we decided to focus on the study of short fragments with different hydropaticity to follow their aggregation. To this aim, we have performed simulations both at physiological salt concentration (0.1 M) and at low ionic strength. The main results of this study have shown that the amino acid composition affects the collagen aggregation. Under physiological conditions the association of less charged regions occurs more easily not requiring a sequential order, whereas highly charged fragments need a sequential assembly. In these conditions, a hydrophobic driving force at the base of the assembly mechanism is proposed, in agreement with Li and co-workers [89], as occurs in protein folding processes. In addition, as for protein folding, also fibrillogenesis is an energy driven process, with a balance of loss of entropy and gain in enthalpy by means of favorable hydrophobic and polar interactions. The similar hydropathic profile of apposing tropocollagen fragments maximizes the number of these favorable contacts. Lowly charged regions could lead the aggregation along the chain in a sequential way, in order to correctly match highly charged regions, whose electrostatic interactions become relevant at short distances.

Several studies have proved that hydroxyproline residue, due to its polarity, might contribute to a correct fibrillogenesis [86] [87]; in particular, it seems to be involved in the formation of water bridges that mediate the tropocollagen approaching. However, we observed that in the final assembly a high number of direct inter-tropocollagens H-bonds are formed by Hyp, involving in particular the carbonyl group of the highly present glycine residue, while water bridges are rarely observed.

Probably, even if water mediation is needed at higher tropocollagen distances, in the final compact assembly direct interactions can be easily formed. This is in agreement with the observation that the fibrillogenesis process occurs with the loss of water from the monomer surface [10], confirmed by the observed decrease in protein-water H-bonds and the specular increase in the inter-tropocollagen H-bonds.

Our simulations at low ionic strength have shown that the electrostatic contribution becomes predominant due to the absence of ions shielding effect. Under this condition, the aggregation of HC fragments is, therefore, favored and fast, and no more sequential but simultaneous. However, the salt bridge formation is now crucial for the association of tropocollagens and, to do this, staggering is not respected. Light scattering experiments performed by Morozova and co-workers have detected that decreasing the ionic strength the process became faster and at low ionic strength collagen molecules form extremely thin microfibers [25]. In light of these considerations, it could be hypothesized that at low ionic force larger fibrillar structures cannot be formed due to the incorrect staggering of tropocollagens.

A parallel study of collagen self-assembly has been performed by UV measurements at three different pH values. Aggregation curves, obtained from UV spectra, allowed to study the different kinetics by varying the pH solutions, showing a non-linear trend. In fact, the slowest process is at pH 7, becoming faster at pH 8 and even faster at pH 5. pH variations slightly below or above neutrality may cause the protonation/deprotonation of some residues, changing the surface charge of the protein. A more rapid process could lead to an incorrect staggering, therefore the pH 7 seems to be the optimal pH value for stability and to obtain the right aggregation. It should be highlighted that *in vivo* collagen self-assembly takes place under physiological conditions (at neutral pH and at 0.1/0.15 M salt concentration). In light of our MD simulations and UV measurements results we could suggest that, under physiological conditions, a slower collagen association, beginning from lowly charged regions and driven by hydrophobic interactions, is needed for the correct packing and subsequent fibrils formation.

As concerns the collagen dehydration study, MD simulations and experimental measurements have been performed in order to investigate on structural changes induced by a dehydration process, because it is known that the water depletion causes irreversible damages in collagen molecules [3] [24].

MD simulations allowed to explore at atomic level the effects induced by dehydration conditions on collagen microfibrils especially with regard to its different hydropathic

profiles. We have analyzed a system where water molecules were depleted up to the first layer around the protein, corresponding to a hydration level of 0.42/0.49 h, and, subsequently, we have studied the rehydration of that system to examine the reversibility of the process. The assembly arrangement in dehydrated state becomes more compact, promoting a more tubular shape and new protein-protein H-bonds, also favored by the lack of outer water. In these conditions, the inner water molecules (i.e. inside the tubular fibril) are preserved and could be relevant in the reversibility of the process, allowing the recovery of the hydrated structural characteristics and avoiding the collapse of the fibril. External water undergoes to rearrangement around the protein, forming clusters preferentially around charged groups. In this conformation, the first layer of water is more anchored to the protein surface than the one in a fully hydrated or rehydrated state.

Rehydration brings back the collagen microfibrils in their initial conditions and water regains its original distribution and mobility. Only a few differences between the simulations of LC and HC models have been observed. Lowly charged fragments tend to maintain the gained tubular shape, even if it should be noted that this conformation was not so different from the hydrated starting structure. Instead, highly charged model recovers completely the flatter shape.

From this MD study, we can conclude that dehydration of a microfibril up to a hydration level of 0.42/0.49 h has shown to be a reversible process. It seems that these dehydration conditions are not a critical threshold for the recovery of the structural features of a fully hydrated fibril. Not by chance, this threshold matches the hydration value over which proteins seem to preserve functionality [73] [19].

A FTIR study has been carried out on collagen molecules to study a whole dehydration-rehydration cycle, reaching strong dehydration levels. Gravimetric measurements quantify the variation of water content during the dehydration-rehydration cycle, spanning the range of hydration levels of 0.6-0.02 h. From the analysis of OH stretching band of FTIR spectra and from the gravimetric measurements, we have obtained sorption isotherms indicating a hysteresis loop in dehydration-rehydration cycle. The hysteresis effect suggests that after the dehydration-rehydration treatment the distribution of water in the collagen sample does not completely correspond to the initial one: charged sites, being required for the nucleation of water clusters, could be buried due to different arrangement of collagen molecules attained during dehydration process. Variations in the structure of collagen sample have been detected by deconvolution of Amide I band, showing an

irreversible increase in the amplitude of beta sheet/turn bands after dehydration-rehydration treatment, as observed in previous FTIR study [3]. This effect can be explained considering a growth of other fibrillar aggregates or an increment of fibril compactness. This result is in agreement with MD simulation results, showing an increase in the fibril compactness under dehydration conditions. It should be highlighted that the lowest hydration value (i.e. 0.02 h) experimentally reached indicates a very strong dehydration compared to the value 0.42/0.49 h obtained in dehydrated systems of MD simulations. In light of these considerations, we can conclude that a moderate dehydration of 0.42/0.49 h, induces reversible changes in the system, whereas strong dehydration, where probably also inner water is lost causes an irreversible collapse of the collagen structure.

References

[1] D. Whitford, "Proteins: Structure and Function," *John Wiley Sons. Ldt, Hoboken, NJ*, 2005.

[2] Nelson and Cox, *Principles of Biochemistry Lehninger*. 2011.

[3] M. G. Bridelli, "Fourier Transform Infrared Spectroscopy in the Study of Hydrated Biological Macromolecules," in *Fourier Transforms - High-Tech Application and Current Trends*, 2017, p. 13.

[4] C. M. Garrett, Reginald H., Grisham, *Biochemistry (5th ed.)*. 2013.

[5] F. W. Kotch, I. A. Guzei, and R. T. Raines, "Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues," *J. Am. Chem. Soc.*, 2008, doi: 10.1021/ja800225k.

[6] M. Yamauchi and M. Sricholpech, "Lysine post-translational modifications of collagen," *Essays Biochem.*, 2012, doi: 10.1042/BSE0520113.

[7] L. Stryer, *Stryer Biochemistry*. 1995.

[8] N. A. Campbell, *Biology*, 3rd ed. 1995.

[9] M. van der Rest and R. Garrone, "Collagen family of proteins.," *FASEB J.*, vol. 5, no. 13, pp. 2814–23, 1991, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1916105.

[10] W. Pompe *et al.*, "Octacalcium phosphate - a metastable mineral phase controls the evolution of scaffold forming proteins," *J. Mater. Chem. B*, 2015, doi: 10.1039/c5tb00673b.

[11] M. Nokelainen, "Characterization of type II collagen expressed in insect cells and production of types I-III collagen in the yeast Pichia pastoris," Oulun yliopisto, Lääketieteellinen tiedekunta, Lääketieteellisen biokemian ja molekyylibiologian laitos, 2001.

[12] D. Fan, A. Takawale, J. Lee, and Z. Kassiri, "Cardiac fibroblasts, fibrosis and extracellular matrix remodeling in heart disease," *Fibrogenesis and Tissue Repair*. 2012, doi: 10.1186/1755-1536-5-15.

[13] J. P. R. O. Orgel, T. C. Irving, A. Miller, and T. J. Wess, "Microfibrillar structure of type I collagen in situ.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 24, p. 9001, 2006, doi: 10.1073/pnas.0502718103.

[14] M. D. Shoulders and R. T. Raines, "Collagen Structure and Stability," *Annu. Rev. Biochem.*, 2009, doi: 10.1146/annurev.biochem.77.032207.120833.

[15] M. Fang and M. M. Banaszak-Holl, “Variation in type I collagen fibril nanomorphology: the significance and origin.,” *Bonekey Rep.*, vol. 2, no. April, p. 394, 2013, doi: 10.1038/bonekey.2013.128.

[16] I. Streeter and N. H. De Leeuw, “A molecular dynamics study of the interprotein interactions in collagen fibrils,” *Soft Matter*, 2011, doi: 10.1039/c0sm01192d.

[17] M. J. Buehler, “Nature designs tough collagen: Explaining the nanostructure of collagen fibrils,” *Proc. Natl. Acad. Sci. U. S. A.*, 2006, doi: 10.1073/pnas.0603216103.

[18] Bryan Stephen Marques, “Reverse Micelle Encapsulation And Its Use In Examining The Interplay Between Hydration And Protein Dynamics,” University of Pennsylvania, 2018.

[19] J. A. Rupley and G. Careri, “Protein Hydration and Function,” *Adv. Protein Chem.*, 1991, doi: 10.1016/S0065-3233(08)60197-7.

[20] W. Wang, “Lyophilization and development of solid protein pharmaceuticals,” *International Journal of Pharmaceutics*. 2000, doi: 10.1016/S0378-5173(00)00423-3.

[21] J. Bella, B. Brodsky, and H. M. Berman, “Hydration structure of a collagen peptide,” *Structure*, vol. 3, no. 9, pp. 893–906, 1995, doi: 10.1016/S0969-2126(01)00224-6.

[22] J. W. Baynes, “The role of AGEs in aging: Causation or correlation,” *Exp. Gerontol.*, vol. 36, no. 9, pp. 1527–1537, 2001, doi: 10.1016/S0531-5565(01)00138-3.

[23] A. J. Bailey, R. G. Paul, and L. Knott, “Mechanisms of maturation and ageing of collagen,” *Mech. Ageing Dev.*, vol. 106, no. 1–2, pp. 1–56, 1998, doi: 10.1016/S0047-6374(98)00119-5.

[24] I. G. Mogilner, G. Ruderman, and J. R. Grigera, “Collagen stability, hydration and native state,” *J. Mol. Graph. Model.*, vol. 21, no. 3, pp. 209–213, 2002, doi: 10.1016/S1093-3263(02)00145-6.

[25] S. Morozova and M. Muthukumar, “Electrostatic effects in collagen fibril formation,” *J. Chem. Phys.*, 2018, doi: 10.1063/1.5036526.

[26] Y. Nomura, M. Yamano, C. Hayakawa, Y. Ishii, and K. Shirai, “Structural Property and in Vitro Self-assembly of Shark Type I Collagen,” *Biosci. Biotechnol. Biochem.*, vol. 61, no. 11, pp. 1919–1923, Jan. 1997, doi: 10.1271/bbb.61.1919.

[27] Z. Xu *et al.*, "Molecular mechanisms for intrafibrillar collagen mineralization in skeletal tissues," *Biomaterials*, 2015, doi: 10.1016/j.biomaterials.2014.10.048.

[28] S. Rele, Y. Song, R. P. Apkarian, Z. Qu, V. P. Conticello, and E. L. Chaikof, "D-periodic collagen-mimetic microfibers," *J. Am. Chem. Soc.*, 2007, doi: 10.1021/ja0758990.

[29] N. Krishnamoorthy, M. H. Yacoub, and S. N. Yaliraki, "A computational modeling approach for enhancing self-assembly and biofunctionalisation of collagen biomimetic peptides," *Biomaterials*, 2011, doi: 10.1016/j.biomaterials.2011.06.074.

[30] P. J. Steinbach and B. R. Brooks, "Protein hydration elucidated by molecular dynamics simulation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 19, pp. 9135–9139, 1993, doi: 10.1073/pnas.90.19.9135.

[31] M. Tarek and D. J. Tobias, "The dynamics of protein hydration water: A quantitative comparison of molecular dynamics simulations and neutron-scattering experiments," *Biophys. J.*, vol. 79, no. 6, pp. 3244–3257, 2000, doi: 10.1016/S0006-3495(00)76557-X.

[32] N. M. Micaêlo and C. M. Soares, "Protein structure and dynamics in ionic liquids. Insights from molecular dynamics simulation studies," *J. Phys. Chem. B*, vol. 112, no. 9, pp. 2566–2572, 2008, doi: 10.1021/jp0766050.

[33] R. Friemann, D. S. D. Larsson, Y. Wang, and D. Van Der Spoel, "Molecular dynamics simulations of a membrane protein-micelle complex in vacuo," *J. Am. Chem. Soc.*, vol. 131, no. 46, pp. 16606–16607, 2009, doi: 10.1021/ja902962y.

[34] A. Lerbret *et al.*, "How strongly does trehalose interact with lysozyme in the solid state? Insights from molecular dynamics simulation and inelastic neutron scattering," *J. Phys. Chem. B*, vol. 116, no. 36, pp. 11103–11116, 2012, doi: 10.1021/jp3058096.

[35] A. Gautieri, S. Vesentini, A. Redaelli, and M. J. Buehler, "Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up," *Nano Lett.*, vol. 11, no. 2, pp. 757–766, 2011, doi: 10.1021/nl103943u.

[36] A. Gautieri, M. I. Pate, S. Vesentini, A. Redaelli, and M. J. Buehler, "Hydration and distance dependence of intermolecular shearing between collagen molecules in a model microfibril," *J. Biomech.*, vol. 45, no. 12, pp.

2079–2083, 2012, doi: 10.1016/j.jbiomech.2012.05.047.

[37] A. Gautieri, S. Vesentini, A. Redaelli, and M. J. Buehler, "Viscoelastic properties of model segments of collagen molecules," *Matrix Biol.*, vol. 31, no. 2, pp. 141–149, 2012, doi: 10.1016/j.matbio.2011.11.005.

[38] D. van der Spoel, E. Lindhal, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, Flexible and Free.," *J. Comp. Chem.*, vol. 26, pp. 1701–1718, 2005.

[39] D. A. Case *et al.*, "The Amber biomolecular simulation programs," *Journal of Computational Chemistry*. 2005, doi: 10.1002/jcc.20290.

[40] B. R. Brooks *et al.*, "CHARMM: The biomolecular simulation program," *J. Comput. Chem.*, 2009, doi: 10.1002/jcc.21287.

[41] J. C. Phillips *et al.*, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*. 2005, doi: 10.1002/jcc.20289.

[42] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *J. Chem. Phys.*, 1993, doi: 10.1063/1.464397.

[43] W. R. P. Scott *et al.*, "The GROMOS biomolecular simulation program package," *J. Phys. Chem. A*, 1999, doi: 10.1021/jp984217f.

[44] C. I. Bayly *et al.*, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, 1995, doi: 10.1021/ja00124a002.

[45] A. D. MacKerell *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, 1998, doi: 10.1021/jp973084f.

[46] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, 1996, doi: 10.1021/ja9621760.

[47] H. Nymeyer and A. E. García, "Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent," *Proc. Natl. Acad. Sci. U. S. A.*, 2003, doi: 10.1073/pnas.2232868100.

[48] C. Tan, L. Yang, and R. Luo, "How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis," *J. Phys. Chem. B*, 2006, doi: 10.1021/jp063479b.

[49] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction Models for Water in Relation to Protein Hydration," 1981.

[50] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *J. Phys. Chem.*, 1987, doi: 10.1021/j100308a038.

[51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, 1983, doi: 10.1063/1.445869.

[52] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, 2000, doi: 10.1063/1.481505.

[53] R. W. Hockney, S. P. Goel, and J. W. Eastwood, "Quiet high-resolution computer models of a plasma," *J. Comput. Phys.*, 1974, doi: 10.1016/0021-9991(74)90010-2.

[54] A. M. Lesk, "Introduction to protein science Architecture, function, and genomics," Second Edi., Oxford University Press, 2010, p. 455.

[55] D. B. Wetlaufer, "Ultraviolet spectra Of Proteins and Amino Acids," *Adv. Protein Chem.*, 1963, doi: 10.1016/S0065-3233(08)60056-X.

[56] U. Gordon C. K. Roberts, Department of Biochemistry University of Leicester, Ed., *Encyclopedia of biophysics*, 2013 Editi. Springer, Berlin, Heidelberg, 2013.

[57] S. R. W. Damien Hall , Ran Zhao, Ian Dehlsen, Nathaniel Bloomfield and J. A. C. Fumio Arisaka, Yuji Goto, "Protein aggregate turbidity: Simulation of turbidity profiles for mixed-aggregation reactions," *Anal. Biochem.*, vol. 498, pp. 78–94, 2016.

[58] A. C. David W. Oxtoby, H. P. Gillis, "Spettroscopia molecolare e fotonica: l'interazione delle molecole con la luce," in *Chimica Moderna*, Quarta., Edises, Ed. Napoli, 2012, p. 1280.

[59] M. Gustafsson, "Spectroscopic Studies of Tissue Using Near-Infrared Raman Microscopy," 1997.

[60] F. Mallamace *et al.*, "Evidence of the existence of the low-density liquid phase in supercooled, confined water," *Proc. Natl. Acad. Sci. U. S. A.*, 2007, doi: 10.1073/pnas.0607138104.

[61] F. Mallamace *et al.*, "Role of the solvent in the dynamical transitions of proteins: The case of the lysozyme-water system," *J. Chem. Phys.*, 2007, doi: 10.1063/1.2757171.

[62] R. Huehne and J. Suehnel, “The Jena Library of Biological Macromolecules - JenaLib,” *Nat. Preced.*, 2009, doi: 10.1038/npre.2009.3114.1.

[63] “The Protein Data Bank,” in *Structural Bioinformatics*, 2005.

[64] EMBL, SIB Swiss Institute of Bioinformatics, and Protein Information Resource (PIR), “UniProt,” in *Nucleic acids research*, 2013, pp. 41: D43-D47.

[65] L. Leo, M. G. Bridelli, and E. Polverini, “Insight on collagen self-assembly mechanisms by coupling molecular dynamics and UV spectroscopy techniques,” *Biophys. Chem.*, 2019, doi: 10.1016/j.bpc.2019.106224.

[66] J. A. Fallas, V. Gauba, and J. D. Hartgerink, “Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions,” *J. Biol. Chem.*, 2009, doi: 10.1074/jbc.M109.014753.

[67] K. Kawahara *et al.*, “Effect of hydration on the stability of the collagen-like triple-helical structure of [4(R)-hydroxyprolyl-4(R)-hydroxyprolylglycine]10,” *Biochemistry*, 2005, doi: 10.1021/bi051619m.

[68] N. Guex and M. C. Peitsch, “SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling,” *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, 1997.

[69] C, D. Petrov, and B. Zagrovic, “Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications.,” *Nucleic Acids Res.*, 2013, doi: 10.1093/nar/gkt416.

[70] C. Margreitter, M. M. Reif, and C. Oostenbrink, “Update on phosphate and charged post-translationally modified amino acid parameters in the GROMOS force field,” *J. Comput. Chem.*, 2017, doi: 10.1002/jcc.24733.

[71] N. Schmid *et al.*, “Definition and testing of the GROMOS force-field versions 54A7 and 54B7,” *Eur. Biophys. J.*, 2011, doi: 10.1007/s00249-011-0700-9.

[72] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” *J Mol Graph*, vol. 14, no. 1, pp. 33–38, 1996.

[73] Roger B. Gregory, “Hydration and Glass Transition Behavior,” in *Protein-solvent interactions*, 1995, p. 192.

[74] D. Mercadante, F. Gräter, and C. Daday, “CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps,” *Biophys. J.*, 2018, doi: 10.1016/j.bpj.2018.01.033.

[75] A. Allouche, “Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares,” *J. Comput. Chem.*, vol.

32, pp. 174–182, 2012, doi: 10.1002/jcc.

[76] R. Gowers *et al.*, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations," *Proc. 15th Python Sci. Conf.*, no. July, pp. 98–105, 2016, doi: 10.25080/majora-629e541a-00e.

[77] R. Araya-Secchi *et al.*, "Characterization of a novel water pocket inside the human Cx26 hemichannel structure," *Biophys. J.*, vol. 107, no. 3, pp. 599–612, 2014, doi: 10.1016/j.bpj.2014.05.037.

[78] R. H. Henchman and J. A. McCammon, "Extracting hydration sites around proteins from explicit water simulations," *J. Comput. Chem.*, vol. 23, no. 9, pp. 861–869, 2002, doi: 10.1002/jcc.10074.

[79] A. Luise, M. Falconi, and A. Desideri, "Molecular dynamics simulation of solvated azurin: Correlation between surface solvent accessibility and water residence times," *Proteins Struct. Funct. Genet.*, vol. 39, no. 1, pp. 56–67, 2000, doi: 10.1002/(SICI)1097-0134(20000401)39:1<56::AID-PROT6>3.0.CO;2-5.

[80] C. Rocchi, A. R. Bizzarri, and S. Cannistraro, "Water residence times around copper plastocyanin: A molecular dynamics simulation approach," *Chem. Phys.*, 1997, doi: 10.1016/S0301-0104(96)00327-8.

[81] R. W. Impey, P. A. Madden, and I. R. McDonald, "Hydration and mobility of ions in solution," *J. Phys. Chem.*, vol. 87, no. 25, pp. 5071–5083, 1983, doi: 10.1021/j150643a008.

[82] A. Debnath, B. Mukherjee, K. G. Ayappa, P. K. Maiti, and S. T. Lin, "Entropy and dynamics of water in hydration layers of a bilayer," *J. Chem. Phys.*, vol. 133, no. 17, 2010, doi: 10.1063/1.3494115.

[83] F. Sterpone, M. Ceccarelli, and M. Marchi, "Dynamics of hydration in hen egg white lysozyme," *J. Mol. Biol.*, vol. 311, no. 2, pp. 409–419, 2001, doi: 10.1006/jmbi.2001.4860.

[84] M. Thuanthong, N. Sirinupong, and W. Youravong, "Triple helical structure of acid-soluble collagen derived from Nile tilapia skin as affected by extraction temperature," *J. Sci. Food Agric.*, 2016, doi: 10.1002/jsfa.7572.

[85] P. Noitup, M. T. Morrissey, and W. Garnjanagoonchorn, "In vitro self-assembly of silver-line grunt type I collagen: Effects of collagen concentrations, pH and temperatures on collagen self-assembly," *J. Food Biochem.*, 2006, doi: 10.1111/j.1745-4514.2006.00081.x.

[86] G. Nemethy and H. A. Scheraga, "Stabilization of Collagen Fibrils by

Hydroxyproline," *Biochemistry*, 1986, doi: 10.1021/bi00359a016.

[87] S. Perret *et al.*, "Unhydroxylated Triple Helical Collagen I Produced in Transgenic Plants Provides New Clues on the Role of Hydroxyproline in Collagen Folding and Fibril Formation," *J. Biol. Chem.*, 2001, doi: 10.1074/jbc.M105507200.

[88] F. H. Silver, "A Molecular Model for Linear and Lateral Growth of Type I Collagen Fibrils," *Top. Catal.*, 1982, doi: 10.1016/S0174-173X(82)80016-2.

[89] Y. Li, A. Asadi, M. R. Monroe, and E. P. Douglas, "pH effects on collagen fibrillogenesis in vitro: Electrostatic interactions and phosphate binding," *Mater. Sci. Eng. C*, 2009, doi: 10.1016/j.msec.2009.01.001.

[90] A. E. Garcia and G. Hummer, "Water penetration and escape in proteins," *Proteins Struct. Funct. Genet.*, 2000, doi: 10.1002/(SICI)1097-0134(20000215)38:3<261::AID-PROT3>3.3.CO;2-H.

[91] M. H. Köhler, R. C. Barbosa, L. B. da Silva, and M. C. Barbosa, "Role of the hydrophobic and hydrophilic sites in the dynamic crossover of the protein-hydration water," *Phys. A Stat. Mech. its Appl.*, vol. 468, pp. 733–739, 2017, doi: 10.1016/j.physa.2016.11.127.

[92] F. Persson, P. Söderhjelm, and B. Halle, "The geometry of protein hydration," *J. Chem. Phys.*, vol. 148, no. 21, 2018, doi: 10.1063/1.5026744.

[93] D. Laage, T. Elsaesser, and J. T. Hynes, "Water Dynamics in the Hydration Shells of Biomolecules," *Chem. Rev.*, vol. 117, no. 16, pp. 10694–10725, 2017, doi: 10.1021/acs.chemrev.6b00765.

[94] G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991.

[95] M. Petukhov, G. Rychkov, L. Firsov, and L. Serrano, "H-bonding in protein hydration revisited," *Protein Sci.*, vol. 13, no. 8, pp. 2120–2129, Aug. 2004, doi: 10.1110/ps.04748404.

Ringraziamenti

Desidero innanzitutto ringraziare la mia tutor Professoressa Eugenia Polverini per la sua grande disponibilità, gentilezza e passione nel seguirmi in questo progetto di dottorato, sempre con ottimi suggerimenti di cui faccio e farò tesoro.

Ringrazio la Professoressa Maria Grazia Bridelli per il suo prezioso aiuto nell'esecuzione delle misure sperimentali e nell'interpretazione dei dati, nonchè per la sua grande disponibilità nel seguirmi in tutto il progetto di ricerca.

Ringrazio la Dottoressa Roberta Bedotti, per il suo aiuto e i suoi consigli sulla preparazione dei campioni sottoposti alle misure UV.

Speciali ringraziamenti vanno ai miei cari e pazienti genitori che hanno da sempre appoggiato e supportato tutte le mie decisioni con mille suggerimenti e idee. È anche, soprattutto, merito loro se ho raggiunto questo traguardo.

Un ringraziamento speciale va al mio dolce Andrea che mi è sempre stato vicino supportandomi e sopportandomi con infinita pazienza e sensibilità, durante questo percorso.

Ringrazio, i miei nonni che, anche se non riesco a vedere spesso, mi sono sempre vicini.

Ringrazio i miei amici con i quali ho condiviso bellissimi momenti durante questi tre anni.

Ringrazio, infine, i miei colleghi e, in particolare, Eleonora, mia vicina d'ufficio, nonchè mia amica, per il suo sostegno e i suoi preziosi consigli.