Insight on collagen self-assembly mechanisms by coupling molecular dynamics and UV spectroscopy techniques

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://ees.elsevier.com

Insight on collagen self-assembly mechanisms by coupling molecular dynamics and UV spectroscopy techniques

Ludovica Leo, Maria Grazia Bridelli, Eugenia Polverini*

Department of Mathematical, Physical and Computer Sciences, University of Parma, Parco Area delle Scienze, 7/A, 43124 Parma, Italy

ARTICLE INFO

Keywords
Tropocollagen self-assembly
Hydroxyproline
Turbidity
Molecular dynamics simulation
Low ionic strength
Hydrophobic driving force

ABSTRACT

Self-assembly of rat tail collagen type I was investigated by means of turbidity measurements and molecular dynamics simulations. Turbidity curves collected at different pH values show that the rate of aggregation was not linear in dependence from pH, with the fastest kinetics at pH 5.0 and the lowest at neutral pH. MD simulations were carried out on two regions with different hydropathicity, monitoring the aggregation of up to four staggered tropocollagen fragments at different ionic strength. At physiological conditions, association of lowly charged regions occurs more easily than for highly charged ones, the latter seeming to aggregate in a sequential way. The first contacts indicate for both regions that the driving force is hydrophobic, the electrostatic contribution becoming relevant at short distance. The direct inter-tropocollagen H-bonds confirm that fibrillogenesis is driven by loss of surface water from the monomers and involves in large percentage hydroxyproline residues. Low ionic strength dynamics leads to the formation of incorrect assemblies, driven by not shielded pairwise charge interactions.

1. Introduction

Collagen is a fibrous protein representing the main constituent of connective tissue in mammals, like skin, bone, tendon and cartilage. The basic structural unit of collagen is tropocollagen formed by 3 intertwined left-handed helices assembled in a triple right-handed helix with shorter terminal domains (so-called telopeptides) of different structure [1]. Tropocollagen is 1.4 nm in diameter and 300 nm in length. The amino-acid sequences of the helices are characterized by the repetition of triplets Gly-Xaa-Yaa, being one-thirds of the X and Y residues either prolines or hydroxyprolines (Hyp) [1]. So far, about 30 distinct collagen types have been identified in vertebrates: in this work type I collagen is considered, which consists of two identical chains (α1) and one different chain (α2) [1]. Collagen triple helices associate in fibrils, where few tropocollagen molecules are staggered side-by-side with a shift of 67 nm between two neighbours tropocollagens, corresponding to a periodicity of 234 residues [5]. Several studies have demonstrated that hydroxyproline induces a triple-helix stabilization through a stereoelectronic effect [2] and it might contribute to a correct fibril formation [3, 4]. Collagen fibrillogenesis is an entropy-driven process by loss of surface water from the monomer [5]. The C-terminal telopeptides are important, but not essential, for initiating proper fibrillogenesis, in fact collagen telopeptides could only accelerate fibril assembly [6]. The supramolecular structures are initially stabilized by non-covalent interactions (e.g. polar and hydrophobic interactions). Later, only after fibrillogenesis, a cross-linking process, involving Lys side chains of the telopeptides, takes place on mature collagen fibrils, to give strength and stability [6].

In this work rat tail collagen self-assembly was studied by coupling molecular dynamics and UV-visible absorption techniques. The different approaches are used to explore different aspects of self-assembly mechanism: the whole process is detected by UV spectroscopy, whereas molecular dynamics enables to focus on molecular interactions, observing the aggregation of up to four staggered collagen fragments. By means of UV technique and the analysis of the turbidity curves, the aggregation process was monitored by measuring the turbidity changes of the solution according to the method of Nomura and co-workers [7]. This method was widely used to study the aggregation of collagen from different species, varying solution properties (like concentrations, pH, temperature) [8–11]. In this work, collagen self-assembly at different pH values was investigated and the rate of aggregation was estimated. Molecular dynamics simulations, on the other side, allowed to investigate how different hydropathic profiles influence collagen self-assembly. Several MD simulations studies on collagen were reported in the literature, in which the intra-fibrillar interactions in an already formed fibril was investigated [12, 13], or in which the aggregation of specifically designed collagen-mimetic peptides was simulated [14, 15]. In this work, a system reproducing two different regions of up to four rat

* Corresponding author.
Email address: eugenia.polverini@unipr.it (E. Polverini)

https://doi.org/10.1016/j.bpc.2019.106224
Received 30 June 2019; Received in revised form 10 July 2019; Accepted 10 July 2019
Available online xxx
0301-4622/

tail collagen staggered fragments was built, and, after putting the fragments at 2 nm distance between each other (not yet in fibrillar disposition) in a box of solvent, their aggregation mechanism was monitored. In addition, for comparison with the experimental work of Morozova and co-workers [16], simulations were performed both at physiological conditions and at low ionic strength.

2. Materials and methods

2.1. UV-visible measurements

2.1.1. Materials

Collagen from rat tail was purchased by Sigma-Aldrich. A collagen solution (1 mg/ml) was prepared by dissolving lyophilized rat tail collagen in acetic acid 0.5 M. Three samples were prepared by dialyzing the solution against phosphate buffer 67 mM at three pH values: pH 5.0, 7.0 and 8.0, at 4 °C for 12 h. Finally the solutions obtained were centrifuged at 10000 *g* for 10 min [9].

2.1.1.1. Measurements In vitro collagen self-assembly was monitored by measuring the turbidity changes of the solutions prepared as described in the previous section. The absorbance increase was recorded at 310 nm, as a function of time. The measurements were performed at 300 K, by means of UV-visible spectrophotometer Jasco 7850, equipped with a thermostat, collecting data every 10 s, according to the method of Nomura et al. [7].

2.2. Computational techniques

2.2.1. Model building of tropocollagen starting structures

Following the experimental conditions, rat tail type I collagen sequence was chosen for MD simulations. Type I collagen is composed by two α1 and one α2 chains. Rat tail sequences were retrieved from Uniprot database [17] (entry: P02454 and P02466).

Due to the big length of the tropocollagen chain (about 1000 amino acids), we chose to separately simulate the aggregation of two kind of fragments with different hydropathicity profiles (Tables 1 and 2), to study the influence of different molecular features on the self-assembly mechanism. Therefore we modeled two 29/30 residues regions (depending on the template length, see below), each made by two α1 and one α2 chains, selected from the rat type I collagen sequence basing on their charges content, and named as "lowly charged" (LC) and "highly charged" (HC) tropocollagen fragments. Subsequently, other fragments were selected from the same sequence with a shift of nx234 residues (with $n = 1$ or 2) upstream and downstream of both LC or HC regions to model the corresponding staggered fragments. Shifted fragments have the same hydropathicity features (Tables 1 and 2).

Even if a rat (*Rattus norvegicus*) tail collagen structure is available in the PDB database [18] (id code: 3HQV [19]), it contains only alpha-carbon atoms at a low resolution (5.16Å) and therefore it is not suitable for MD simulation. Therefore we built a model for each triple helix starting structure, basing on the following pdb templates: 1WZB [20], that is rich of POG triplets, for the lowly charged fragment and its staggered sequences, and 2KLW [21] in which several charged amino acids are present, for the highly charged fragment and its staggered sequences. The tropocollagen models were built with Swiss-model server, after a deep inspection of the alignments [22].

For all tropocollagens, Vienna PTM 2.0 server [23,24] was used to transform selected prolines and lysines in hydroxyprolines and hydroxylysines, as indicated in the 3HQV structure.

Percentages of residue type for LC and HC tropocollagen models are reported in Table 2. Fragments are indexed from T0 to T3, according to their positions in the whole rat tail collagen sequence, from N-terminal to C-terminal.

2.2.2. MD simulations

To study collagen monomers assembly, MD simulations were performed using GROMACS software package [25] with the Gromos96 ffG54a7 force-field [26].

Tropocollagen fragments were initially placed at 2.0 nm distance between each other using VMD software [27], and different simulations with two, three and four tropocollagens with staggered sequences were performed for LC and HC models. Gromos96 ffG54a7 force-field [26] was built up inserting modified residues topology (hydroxyproline and hydroxylysine) found in Vienna PTM 2.0 server [23,24].

In each simulation, N- and C-termini were kept in the neutral form, as well as histidine residues. Two, three or four tropocollagen fragments were inserted in a cubic box of water ($10 \times 10 \times 10$ nm for LC models and $11 \times 11 \times 11$ nm for HC models) with periodic boundary conditions, adding enough Na and Cl ions to neutralize the system and to reach a physiological concentration (0.1 M). For each simulation, NTP ensemble was used, with Berendsen barostat for pressure coupling and V-rescale thermostat for temperature coupling. PME summation and potential switch methods were used to calculate electrostatic contributions and Van der Waals interactions, respectively. Before full MD simulation, minimization of the whole system was carried out, followed by a 100 ps position restrained MD simulation to relax solvent molecules around the protein.

Full MD simulations 200 or 300 ns long (depending on the RMSD profile, that monitors the reached stability) were performed at 300 K and 1 atm, with an integration step of 2 fs. Each simulation was replicated at least three times.

In the same conditions, simulations with three and four tropocollagen were performed also with very low ionic strength: the system was

Table 1
Sequences of staggered fragments in LC and HC models. The bold residues are hydroxylated.

Tropocollagen of LC model	α1 chain	α2 chain
T0	A**P**GIAGA**P**GF**P**GARGPSGPQGPSGA**P**GPKG	L**P**GVAGA**P**GL**P**GPRGI**P**GPVGAAGATGPRG
T1	PAGS**P**GFQGL**P**GPAGP**P**GEAGK**P**GEQGV**P**G	PAGP**P**GFQGL**P**GPSGTAGEVGK**P**GERGLPG
T2	P**P**GATGF**P**GAAGRVG**P**PGPSGNAGP**P**GP**P**G	PPGMTGF**P**GAAGRTG**P**PGPSGITGP**P**GP**P**G
T3	DRGI**K**GHRGFSGLQGP**P**GS**P**GS**P**GEQGPSG	ARGLPGLKGHNGLQGLPGLAGLHGDQGAPG
Tropocollagen of HC model	**α1 chain**	**α2 chain**
T0	M**K**GHRGFSGLDGAKGDTGPAGPKGE**P**GS**P**G	F**K**GIRGHNGLDGLKGQ**P**GAQGV**K**GE**P**GA**P**G
T1	ADGVAGPKGPAGERGS**P**GPAGPKGS**P**GEAG	ADGRAGVMGPPGNRGSTGPAGVRGPNGDAG
T2	ERGAAGL**P**GPKGDRGDAGPKGADGSPGKDG	ERGAAGI**P**GGKGEKGETGLRGEIGNPGRDG
T3	QRGERGF**P**GL**P**GPSGE**P**GKQGPSGASGERG	SRGERGQ**P**GIAGALGE**P**GPLGIAGP**P**GARG

Table 2
Percentages of residue type in LC and HC models.

LC model				
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%
HC model				
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%

only neutralized adding Cl ions. Each simulation was replicated two times.

The trajectories analysis was performed with the GROMACS subroutines [25] and with the VMD software package [27]. The distance vs. time between two tropocollagens and the inter-tropocollagen hydrogen bonds (with a 3.5 Å cut-off distance and 30° cut-off angle) were calculated with the GROMACS subroutines [25]. Distance between tropocollagens was monitored computing the distance between their center of mass (COM).

Contact and interaction maps analysis was carried out by Conan tool [28]. The distance between two residues is defined as the shortest distance between any two heavy atoms of the residues. The maps of type of interactions between pair of tropocollagens were built choosing the frame in which the first stable contacts occurred producing a minimum of fifteen points.

3. Results

3.1. UV measurements: kinetics of collagen self-assembly

Fig. 1A shows the turbidity curves recorded measuring the UV absorbance of the solution at different pH values. The curves were normalized to the maximum absorbance to display the aggregate fraction (AF) vs time.

All curves show a sigmoid profile, with three distinct phases (shown in Fig. 1B for pH = 7.0) typical for filamentous protein aggregation according with literature [8].

Lag phase is the first step in which the absorbance does not change: the fibril nucleation starts. Growth phase is the second step, in which the absorbance increases, because the cores of fibrils grow. Plateau is the last step in which the absorbance tends towards a constant value: the formation of three-dimensional networks of fibrils takes place. From the recorded curves it is possible to estimate the half-time (t_{50}) and the lag time (t_{lag}). The t_{50} is the time at which half of the aggregates are formed. The lag time, obtained by the intercept on the abscissas of the tangent at time (t') in which the growth rate slope is at its maximum value (V_{max}), corresponding to the derivative of the curve at inflection point. Table 3 lists V_{max}, t_{lag} and t_{50} values measured for each curve recorded at different pH.

The data show that the collagen aggregation parameters are critically dependent on the pH conditions. The inset of Fig. 1A enlarges the view of the first 35 min to highlight the differences in the lag phases. The self-assembly process at pH 5.0 has the shortest lag phase and the fastest growth phase: half aggregate fraction is formed in just 23 min, almost a quarter of the time taken at pH 7.0. On the other hand, collagen aggregation at pH 7.0 has the longest lag phase and the slowest growth phase, revealing that at neutral pH collagen aggregates more slowly than at the other two pHs. The kinetics at pH 8.0, in fact, shows shorter values than pH 7.0, indicating that the parameters variation is not linear, with the maximum values reached at physiological pH.

3.2. MD simulations

3.2.1. Simulations at physiological conditions

The tropocollagen self-assembly for both LC and HC models, was monitored by means of MD simulations 200/300 ns long on two, three and four staggered tropocollagen fragments, inserted parallel at 2 nm distance in a water box. Each simulation was repeated three times.

3.2.1.1. Lowly charged model As it can be observed in the plot of the COM distance (Fig. 2), aggregation of two and three tropocollagens is fast (< 15 ns), and is observed in all the three replicas. Instead, the complete assembly of four tropocollagens was seen only in two of four replicas, while in the other cases the tropocollagens get close but remain trapped in an only partially aggregated conformation.The final arrangement of three tropocollagens (Fig. 3A) is almost planar. The COM distance of T1-T3, in fact, remains higher than the other two (Fig. 2B).Conversely, the four tropocollagen aggregate in a quite tubular shape (Fig. 4A), and this arrangement is obtained after the aggregation in two pairs, even if the first contacts between the four fragments occur almost at the same time, as can be observed also from the time scale of the interaction maps in Fig. 4C. Then, the four tropocollagens start forming pairs in about 10 ns, and subsequently the pairs aggregate

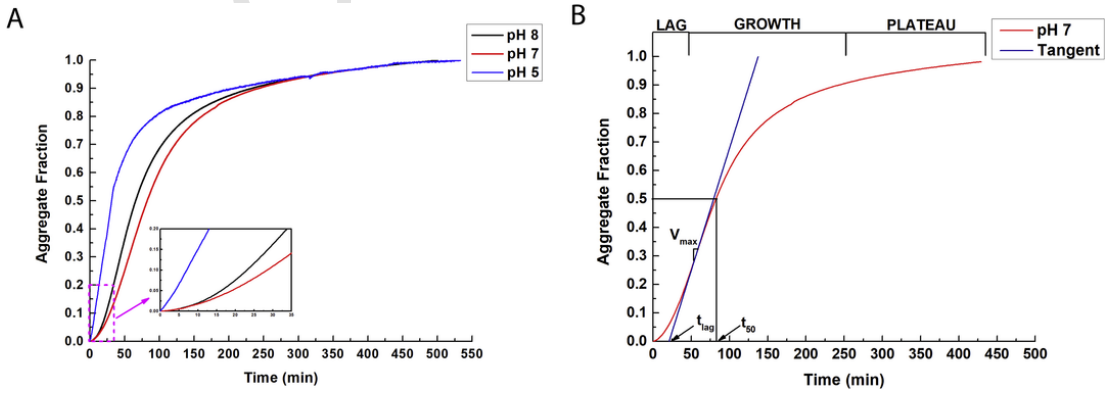

Fig. 1. Turbidity curves. (A) Aggregate fraction versus time at three different pH conditions. In the inset, the view of the first 35 min is enlarged to highlight the differences in the lag phases. (B) Aggregate fraction versus time at pH 7.0. Three distinct phases are recognized: lag, growth and plateau. Lag time is determined by the intercept of the tangent at t' (in which the rate is V_{max}) on the time axis .

Table 3
V_{max}, t_{lag} and t_{50} by varying pH.

pH	V_{max} (aggregate/min)	t_{lag} (min)	t_{50} (min)
pH 5.0	0.030 ± 0.001	3.5 ± 0.1	23.5 ± 0.1
pH 7.0	0.008 ± 0.001	21.0 ± 0.1	82.3 ± 0.1
pH 8.0	0.010 ± 0.001	14.9 ± 0.1	65.5 ± 0.1

forming the final protofibril in about 55 ns. Such a behavior is evident observing the inter-tropocollagen COM distance plots of Fig. 2C, in which T0-T1 and T2-T3 pairs reach their minimum COM distance in less than 10 ns, while T1 and T2 reaches the same distance only after about 55 ns. The first variable part of the plot reflects the oscillations of the two formed dimers before reaching the final conformation.Contact and interaction maps help to visualize which tropocollagens come closest and when, and which type of interactions come in the first events of a simulation. Results for simulations with three and four tropocollagens are shown in Figs. 3 and 4.In the three tropocollagens assembly the contacts between T1 and T3 are almost absent (Fig. 3B), in agreement with a nearly planar aggregate. On the contrary, the four fragments map (Fig. 4B) reflects a quite tubular arrangement.In both simulations, triple helices T1 and T3 come near T2 almost simultaneously. From the maps reporting the kind of interactions between two tropocollagens at

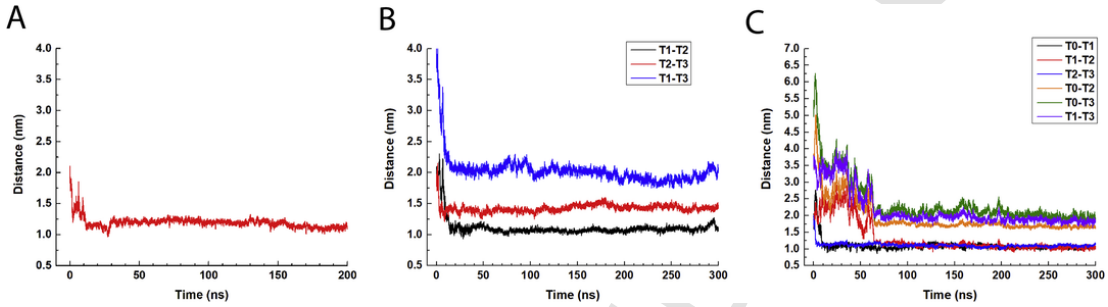

Fig. 2. Inter-tropocollagen COM distance of LC fragments. Plots of inter-tropocollagen COM distance vs. time in the simulations with two (A), three (B) and four (C) LC tropocollagens.

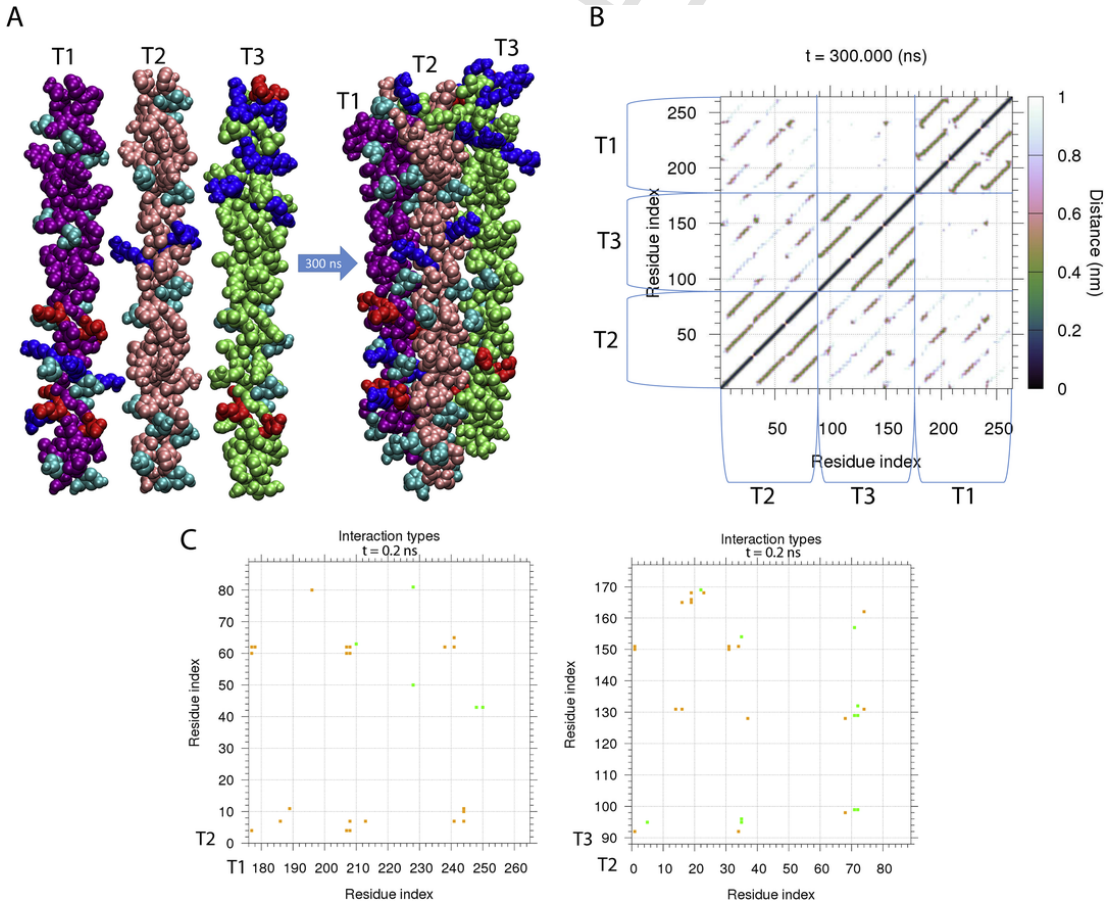

Fig. 3. Association of three LC tropocollagens. (A) Van der Waals representation of the systems with three LC tropocollagens before and after 300 ns MD simulations. Tropocollagen fragments are indexed from T1 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal, and colored in different way. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring: acidic residues are red, basic residues blue and Hyp cyan. (B) Contact map of the last structure (300 ns) of the trajectory. (C) Interactions maps of first contacts in T1-T2 and T2-T3 pairs: hydrophobic type are indicated with orange spots, polar with green spots and salt bridges in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

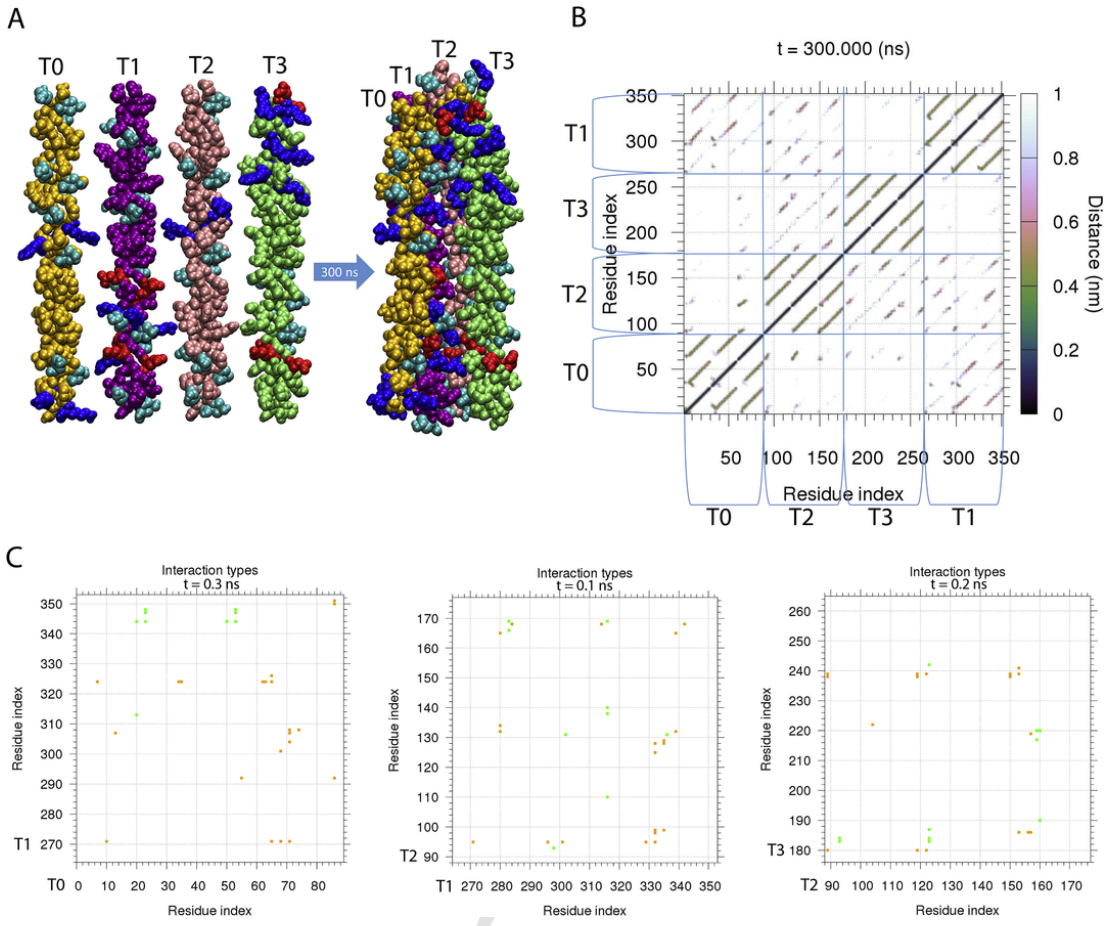

Fig. 4. Association of four LC tropocollagens. (A) Van der Waals representation of the systems with four LC tropocollagens before and after 300 ns MD simulations. Tropocollagen fragments are indexed from T0 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal, and colored in different way. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact map of the last structure (300 ns) of the trajectory. (C) Interactions maps of first contacts in T0-T1, T1-T2 and T2-T3 pairs; the different interaction types are colored as in Fig. 3.

the first contacts (Figs. 3C, and 4C), it is worthwhile to note that they are mainly hydrophobic (orange spots).The number of inter-tropocollagen hydrogen bonds formed between four fragments is reported in Fig. 5A. It is well known that the absolute number of H-bonds depends on the threshold (in distance and angle) used to detect them; however, the reported trend is obtained also changing the threshold values and what is important here is the comparison with the other profiles calculated in the same conditions.The oscillations in the first part of the plot reflect the dynamics of a pair coupling, already observed in the distance plot.In the same Fig. 5A, also the contribution of hydroxyproline residues only are reported. In fact, LC fragments are rich in hydroxyprolines, a putative key residue in the tropocollagen assembly and fib-

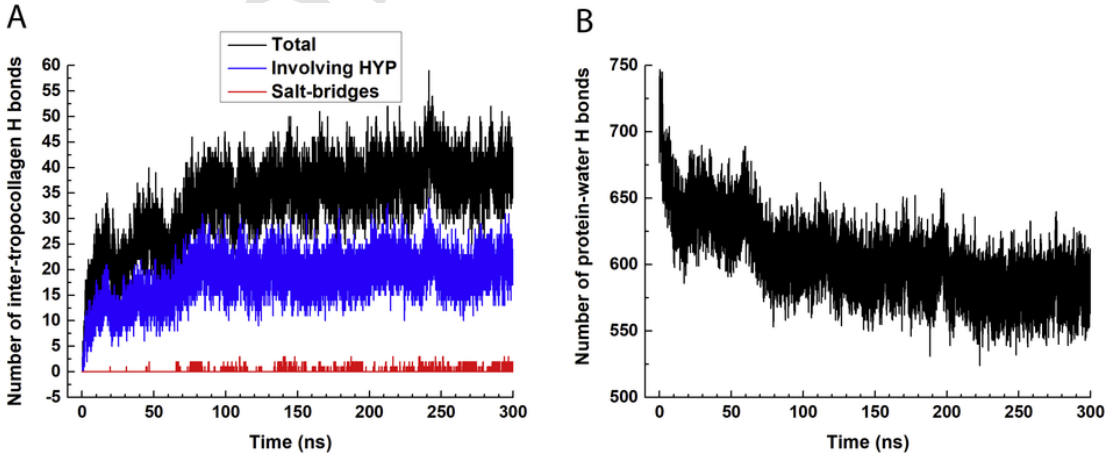

Fig. 5. Inter-tropocollagen and protein-water hydrogen bonds in four LC fragments. (A) Plot of inter-tropocollagen hydrogen bonds vs. time in the simulations with four LC triple helices. In each plot hydrogen bonds involving hydroxyproline residues only (blue line) and salt bridges (red line) are also reported. (B) Number of protein-water hydrogen bonds versus time in the simulations with four LC triple helices. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ril formation [3, 4]. In particular, in T0, T1 and T2 fragments, the number of hydroxyprolines is between 14% and 16% of the total number of residues, while in T3 is 7% (Table 2). In the simulation with four tropocollagens, the final contribution of hydroxyprolines is about 50% of the total number of H-bonds, that is very high considering that the percentage of the total Hyp residues present in the four tropocollagen is about 11%. Mainly, the side chain of Hyp makes a direct H-bond with the glycine backbone carbonyl oxygen, while H-bonds mediated by water are rarely observed in the final assembly.The number of inter-tropocollagen salt bridges formed in LC model and reported in the same figure is very low, due to the low number of charged residues and their disposition along the sequences.Hydrogen bonds formation between protein and water was also monitored (Fig. 5B). The number of protein-water hydrogen bonds decreases as tropocollagens are getting close. As expected, the profile is specular to the one of the inter-tropocollagen H-bonds formation.

3.2.1.2. Highly charged model Differently from lowly charged model, while the assembly of two tropocollagens was observed in all three replicas, three and four tropocollagens were not able to aggregate in a simulation time of 300 ns. It was observed [29] that the rotational freedom of such short fragments is rate limiting for lateral association. Therefore it was tried to add the third and/or fourth tropocollagens to a previously aggregated pair, putting each one at a starting distance of 2.0 nm (Figs. 6A and 7A).For three fragments, two kind of starting structure were built: T3 near the couple T1-T2 and T1 near the couple T2-T3. For each simulation two or three replicas were run. Starting from T3 near T1-T2, the aggregation was observed only one in three cases, with T3 arranged diagonally near the other two fragments (data not shown). Instead, starting from T1 near T2-T3 the assembly occurs in both two replicas (Fig. 6A).In light of these results, the starting structure for four tropocollagens was built using T2-T3 as the assembled pair. In this conditions, we were able to see only one aggregate in three replicas (Fig. 7A).Observing the COM distance plots (Fig. 8), the three tropocollagens assembly seems more compact than in the PC model, with the T1-T2 distance at the same value as the previously assembled T2-T3, as it was easier to get close to already assembled fragments.The aggregation of four tropocollagens in the HC model occurs in a more sequential way, reaching, however, a tubular shape. As highlighted in the inset of Fig. 8 B, firstly T1 comes near T2, secondly T0 get close T1 forming the couple T0-T1 and finally the two couple of fragments T0-T1 and T2-T3 aggregate.The contact maps of the last structure indicate also in the HC model an almost planar arrangement for the three tropocollagens assembly and a tubular one for the four fragments (Figs. 6B and 7B).Even if in the first contacts (Figs. 6C, and 7C) there are more polar interactions (green and purple spots) than in the lowly charged model, the hydrophobic contribute is relevant, especially in pair T1-T2 that possess the highest percentage of hydrophobic residues and that are the fragments that move close faster.The number of inter-tropocollagen hydrogen bonds is reported in Fig. 9 for the four fragments dynamics. The absolute value is higher than that of LC model, but it is proportional to the increment in number of polar residues (comprehending charged ones) that can form H-bonds (about 35% for the PC model, and about 40% for HC one).As expected, due to the higher number of charged residues (23.8% of the total, on average,

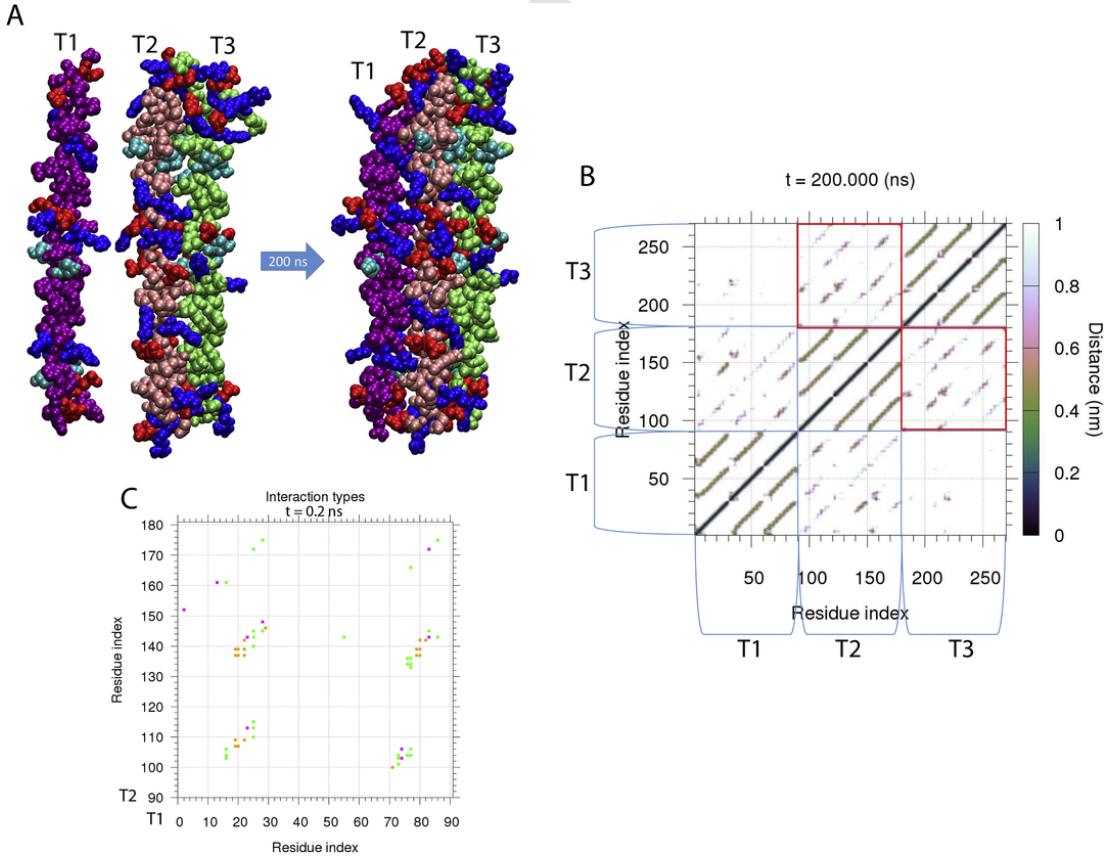

Fig. 6. Association of three HC tropocollagens. (A) Van der Waals representation of the systems with three tropocollagens before and after 200 ns MD simulations. Tropocollagen fragments are indexed from T1 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact map of the last structure (200 ns) of the simulation starting with T1 near T2-T3 pair. The red squares highlight the T2-T3 pair, already in contact in the starting structure. (C) Interactions map of first contacts in T1-T2 pair; the different interaction types are colored as in Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

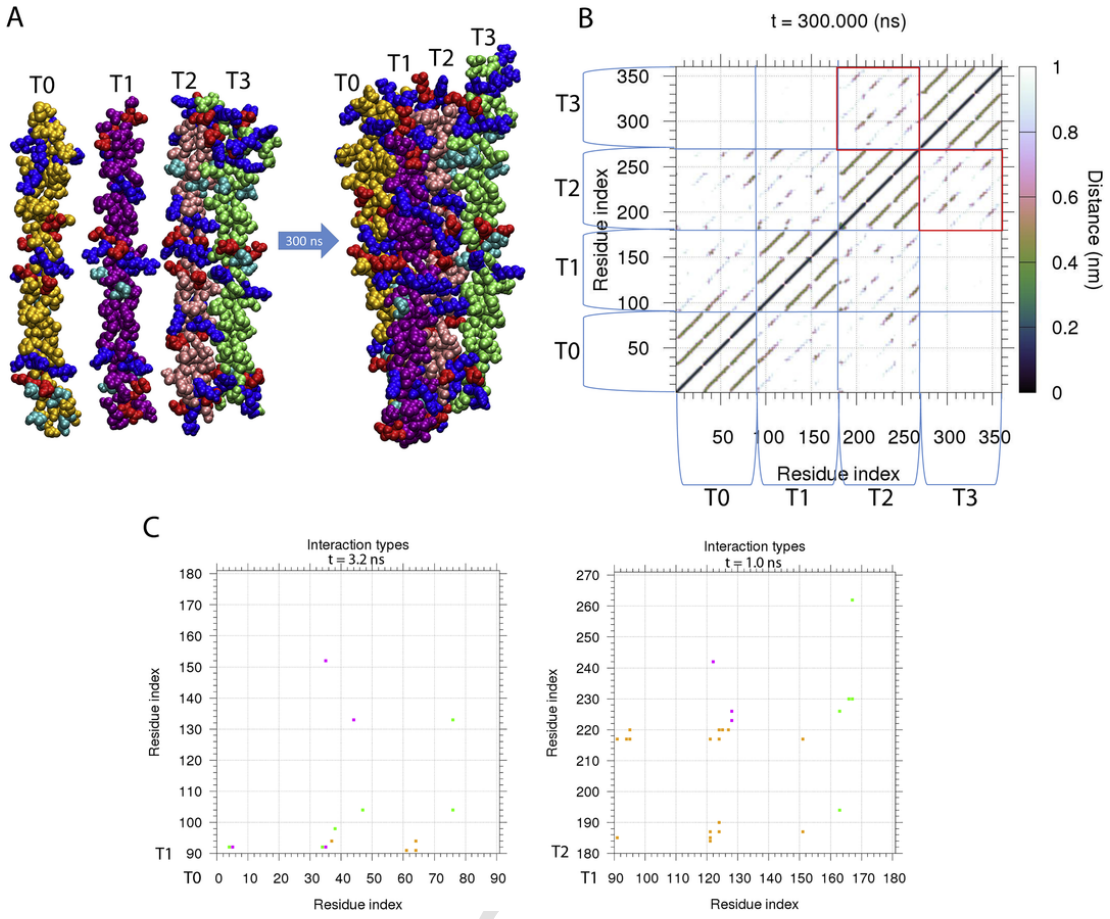

Fig. 7. Association of four HC tropocollagens. (A) Van der Waals representation of the systems with four HC tropocollagens before and after 300 ns MD simulations. Tropocollagen fragments are indexed from T0 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact maps of the last structure (300 ns) of the simulation starting with T2-T3 pair already associated. The red squares highlight the T2-T3 pair. (C) Interactions maps of first contacts in T0-T1 and T1-T2 pairs; the different interaction types are colored as in Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

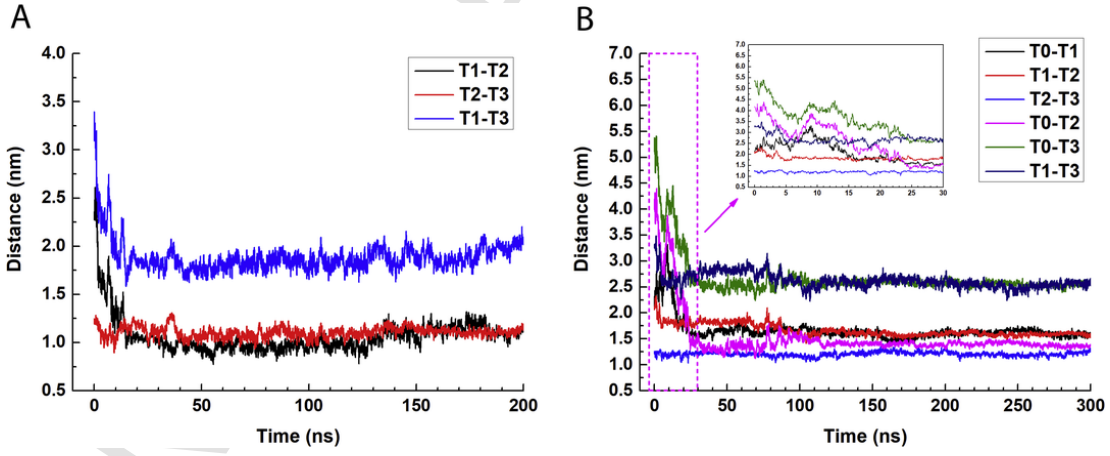

Fig. 8. Inter-tropocollagen COM distance of HC fragments. Inter-tropocollagen distance vs. time in the simulations with three (A) and four (B) HC triple helices; in the inset, the view of the first 30 ns is enlarged. The COM distance between two HC tropocollagen has the same profile as for LC fragments and therefore was not reported.

vs. 9.1% of LC model), the number of salt bridges is much more consistent than in the LC model.Even if the total number of hydroxyproline residues in the four HC tropocollagens is lower (about 6% of the total residues), their contribution to the H-bond formation is 20% of the total, that is still high compared to the LC model.

3.2.2. Simulations at low ionic strength

MD simulations of tropocollagen self-assembly were performed for both LC and HC models also at low ionic strength, i.e. adding the only chloride ions to keep the system neutral. The simulations were performed on systems with three and four tropocollagens and each simulation was repeated two times.

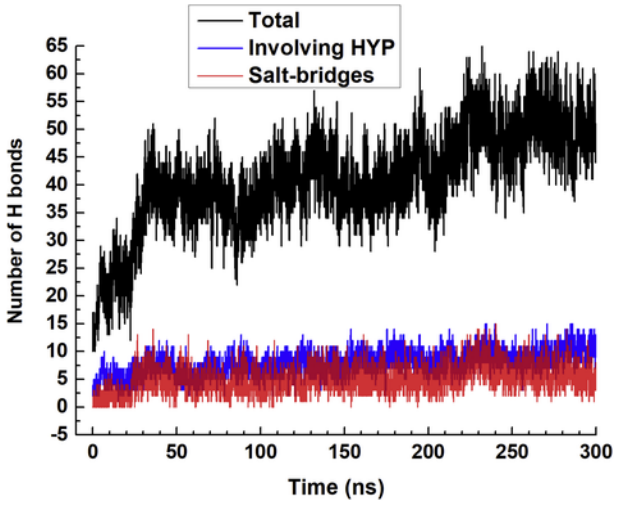

Fig. 9. Inter-tropocollagen hydrogen bonds in four HC fragments. Plot of inter-tropocollagen hydrogen bonds vs. time in the simulations with four HC triple helices. In each plot hydrogen bonds involving hydroxyproline residues only (blue line) and salt bridges (red line) are also reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.2.1. Lowly charged model At low charge conditions, aggregation of tropocollagen is very difficult to occur. In both replicas with three tropocollagens T1-T2 is the only pair formed, while T3 was not able to aggregate correctly (Fig. 10). Association of four tropocollagens was seen only in one cases (Fig. 11A), because in the other one tropocollagens is able to form only pairs.By the contact maps of four tropocollagens (Fig. 11B), a parallel arrangement is evident, no more tubular:

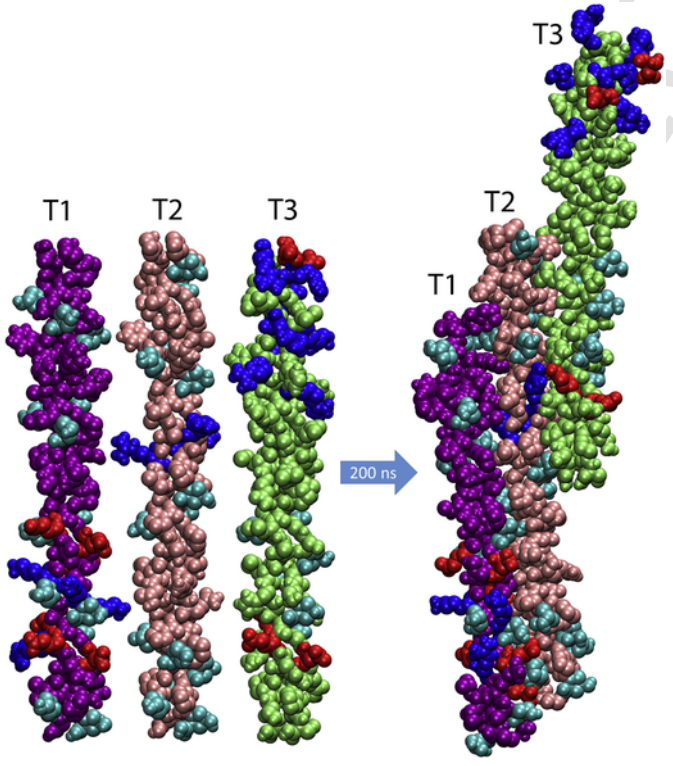

Fig. 10. Association of three LC tropocollagens at low ionic strength. Van der Waals representation of the systems with three LC tropocollagens before and after 200 ns MD simulations at low ionic strength. Tropocollagen fragments are indexed from T1 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3).

contacts are present only between T0-T1, T1-T2 and T2-T3 pairs. At level of the first interactions, small regions of the triple helices move close almost simultaneously, but the whole tropocollagens aggregate sequentially: first the pair T2-T3, then the pair T1-T2 and finally the pair T0-T1. The first interactions are mainly hydrophobic, even if the polar contribution is also present in particular between T2 and T3, which aggregation occurs first and involve the whole chains almost contemporarily (Fig. 11C).

3.2.2.2. Highly charged model The behavior of HC models at low ionic strength is quite different: the three tropocollagens aggregation (Fig. 12A) occurs in both replica simulations, and it is possible to observe the aggregation of three and/or four separated tropocollagens without the previous pair assembly needed at physiological conditions. However, union of four tropocollagens was seen only one in two cases, and the fragments are shifted to put in contact the charged residues, in this way losing the correct staggering (Fig. 13A). The first contacts, in fact, are here for the first time mainly polar, both for three and for four fragments aggregation, with a high presence of salt bridges (Figs. 12C and 13C).The two contact maps (Figs. 12B and 13B) show the almost planar arrangement of the tropocollagens. In the four fragment assembly, the absence of contacts in the upper left or lower right corners of T0-T1, T1-T2 and T2-T3 grid squares indicates the shift of the chains.

4. Discussion and conclusions

4.1. Assembly mechanism in physiological conditions

The choice of investigating two regions of the tropocollagen with different hydropathicity features was useful to study the differences in the aggregation mechanism due to the different amino acid composition. It was already hypothesized, in fact, [3] that the specific interaction of the side chains, as well as their size, could affect the packing of tropocollagens. Our choice was corroborated by the observation that, along the whole sequence, with a glance to the crystal structure [19], staggered fragments have similar hydropathic features, with the charged regions forming a sort of blocks along the fibril.

As it was already observed [12], fibrillogenesis resembles to the process of protein folding, in particular regarding two aspects: the amino acidic sequence code determines the final three-dimensional structure of the system, and folding, as well as fibrillogenesis, is an energy driven process in which entropy loss is balanced by enthalpic gain obtained by favourable hydrophilic and hydrophobic interactions. From this point of view, the staggering of the tropocollagens sequences assures the best inter-tropocollagen interaction maximizing the number of favourable contacts, apposing sequence with similar hydropathicity profiles.

PC and HC models, with their different amino acidic composition, show different ability to associate. MD results indicate that the less charged regions assembly more easily at physiological conditions, and come in contact with no sequential order. On the contrary, for the highly charged model in the same conditions, sequential aggregation seems to be the only possible way: the binding of two tropocollagens promotes following bindings, and it seems to be sequence-dependent.

Looking at the first contacts formed among the tropocollagens, both LC and HC fragments exhibit mainly hydrophobic interactions, indicating a hydrophobic driving force for the tropocollagen assembly mechanism, in agreement with the suggestions of Li and co-workers [30]. This evidence could be reinforced by the analogy with the protein folding process, which is mainly driven by hydrophobic effects, and in agreement with the low ionic strength results discussed below. Considering the assembly of a whole fibril, we could hypothesize a hydrophobic driven aggregation starting from less charged regions that could lead the subsequent association of the rest of the chain. However, a sequential aggregation could be required, to correctly match the highly

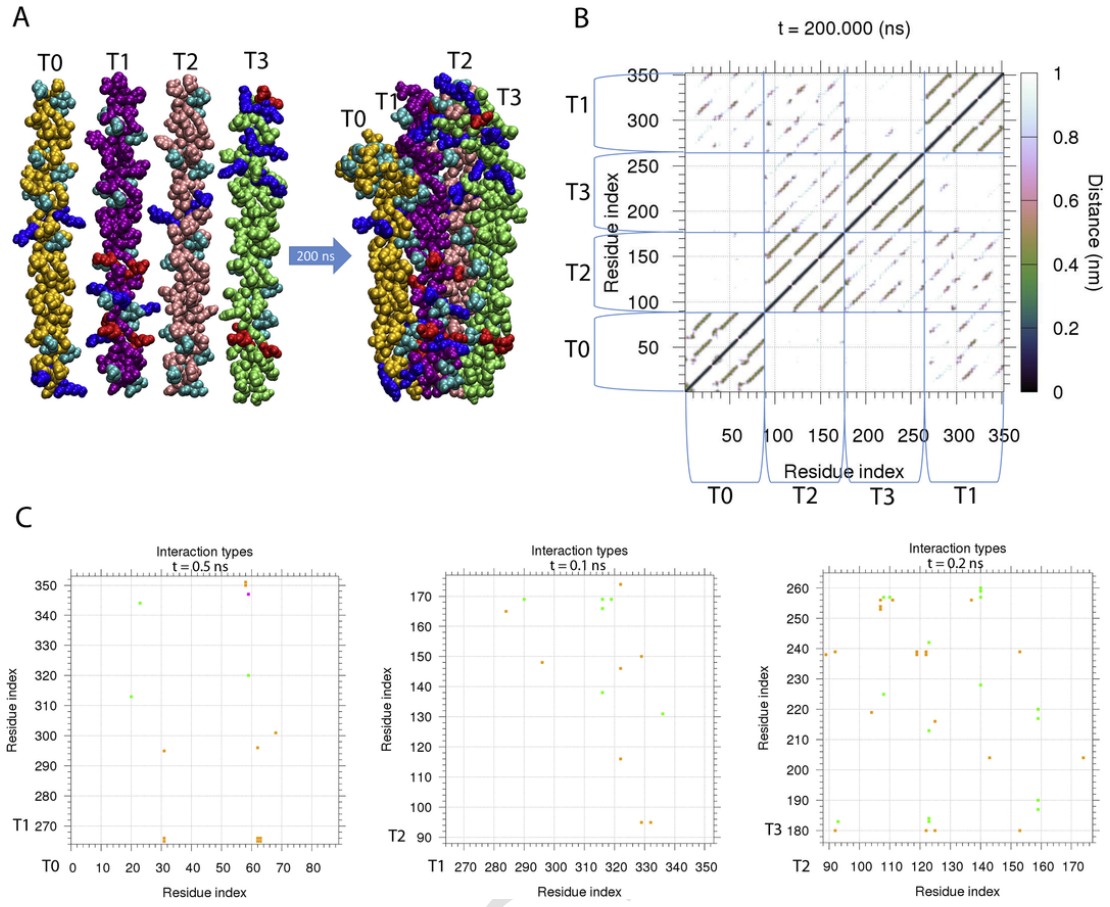

Fig. 11. Association of four LC tropocollagens at low ionic strength. (A) Van der Waals representation of the systems with four LC tropocollagens before and after 200 ns MD simulations at low ionic strength. Tropocollagen fragments are indexed from T0 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact map of the last structure (200 ns) of the trajectory. (C) Interactions maps of first contacts in T0-T1, T1-T2 and T2-T3 pairs; the different interaction types are colored as in Fig. 3.

charged regions, being the electrostatic contribution relevant in the short range distances [30].

As indicated by MD simulations, the arrangement of three tropocollagens is mostly planar, and the subsequent adding of a fourth tropocollagen lead to a more tubular shape, in agreement with what we can argue from the fibril crystal structure. Specular to the increase of the inter-tropocollagen H-bonds with the formation of the assembly, the number of protein-water hydrogen bonds decreases as tropocollagens are getting close, confirming that fibrillogenesis is driven by loss of surface water from the monomers [5].

It was already demonstrated that hydroxyproline plays a role in the stability of collagen, both intra- and inter-triple helix [2, 3]. Polarity of Hyp improves the complementarity of the staggered tropocollagens [3]. Our results show that hydroxyprolines make a high number of direct inter-tropocollagens H-bonds, in particular with carbonyl group of glycine, a highly present residue. Instead, water bridges involving Hyp, that are reported to mediate the inter-tropocollagen approaching [12], are here rarely observed in the final assemblies. Even if they are formed during the fragments approaching, the length of the simulation permits to obtain a quite compact structure, that in fact in these regions is tightly packed, in which the direct inter-protein interaction can be easily formed without the mediation of water molecules, necessary to span greater distance.

4.2. Simulations at physiological condition versus simulations at low ionic strength

Lowly charged and highly charged models act in a different way in dependence on ionic concentration.

At low ionic strength the tropocollagens associate only when they are able to bring the charged residues close. The electrostatic interaction, in fact, are enhanced, due to the loss of ions shielding effects. Therefore, the salt bridge formation is now the driving force for aggregation: actually, it is strictly needed to associate. The fragments tend to reach a fast, complete coupling of the oppositely charged residues, even if it leads to shifted assemblies. From this point of view, in LC model, the low ionic strength seems to penalise self-assembly, due to the low number of charged residues and, in some cases, to the loss of charges coupling. For example, the association of three LC tropocollagens is obtained only when a wide shift of the last fragment along its axes occurs (Fig. 10), and Asp/Glu25 (α1/α2 chain) has moved to interact with the oppositely charged Arg13. The absence of ionic shielding, in fact, penalizes in particular negatively charged residues, because the positive ones are partially shielded by chloride ions put in the solvent to neutralize the net charge of the system.

Following these considerations, at low ionic strength conditions the assembly of HC fragments should be favoured, and in fact it is, with the single triple helices aggregating simultaneously, not sequentially like at physiological conditions. The association is faster, contrary to the behavior at physiological salt concentration, where ions slow down the

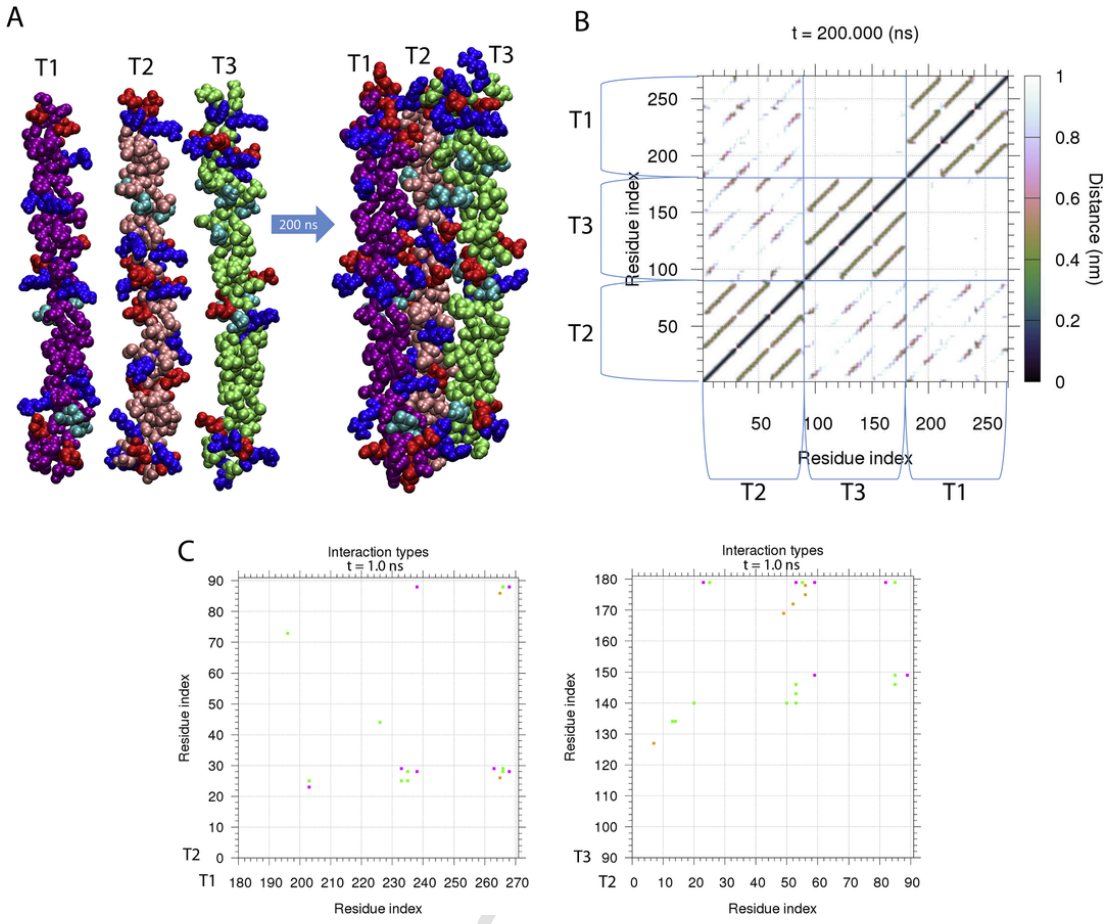

Fig. 12. Association of three HC tropocollagens at low ionic strength. (A) Van der Waals representation of the systems with three HC tropocollagens after 200 ns MD simulations at low ionic strength. Tropocollagen fragments are indexed from T1 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact map of the last structure (200 ns) of the trajectory. (C) Interactions plots of first contacts in T1-T2 and T2-T3 pairs; the different interaction types are colored as in Fig. 3.

pairwise charge interactions and the aggregation of HC fragments can occur only at short distance, after displacement of oppositely charged ions around the residues (see Fig. 14 as an example).

Despite of the favourable conditions created by low ionic strength, however, also the HC fragments assembly occurs with a shift of tropocollagen along their axes, forced by the spatial coupling of oppositely charged residues. As it can be observed also from the fibril crystal structure, the staggering of the tropocollagens must be strictly respected, therefore suggesting that this shifted assemblies are uncorrected. In agreement with this hypothesis, it was experimentally observed that in low ionic strength solutions collagen forms extremely thin microfibers, and the fibrillization process is fast [16]. We can therefore hypothesize that in low ionic strength conditions, the uncorrected packed tropocollagens are not able to form larger fibrillar structures.

It should be underlined that in vivo collagen self-assembly occurs at physiological ionic strength of about 0.1/0.15 M, where the results indicate an association that starts from lowly charged regions, is driven by hydrophobic interactions, and permits the correct packing.

4.3. Assembly mechanism depends on pH

Curves obtained from UV measurements show different aggregation kinetics by varying the pH of the solution. The behavior shows a non linear trend, with the curve at pH 7.0 having the slowest fibrillogenesis process, with the longest nucleation phase and the slowest grow phase. The kinetics is faster at pH 8.0 and even faster at pH 5.0. These difference may be caused by changes in the charge of the protein. It was demonstrated that at extreme pH values, fibrils cannot be formed, due to the repulsive effects of the highest net charge [16]. At pH values slightly above or below neutrality, only some charged residues start to become protonated, decreasing the surface charge of the protein. If we follow our hypothesis of an aggregation more favourable for less charged regions, and probably starting from them, we could partially explain why the association process is faster at pH 5.0 and pH 8.0 than at pH 7.0. Deeper investigation at molecular lever are planned for a future work, to better understand the pH dependent changes in the chain interactions that influence the fibrillogenesis.

Acknowledgements

This work was supported by Local Funding for Research (FIL funds), University of Parma, to EP and MGB, and benefits from the HPC (High Performance Computing) facility of the University of Parma, Italy. We thank Dr. Roberta Bedotti, University of Parma, for kind assistance in samples preparation.

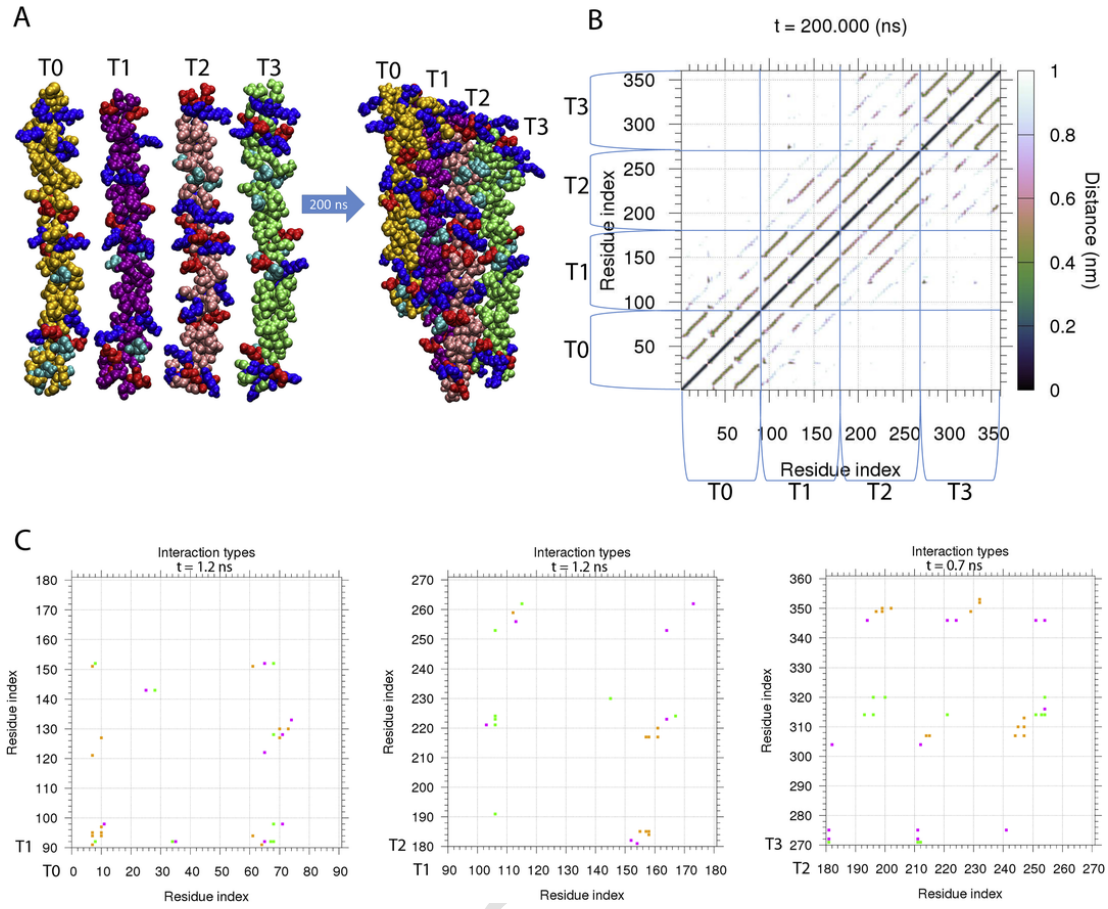

Fig. 13. Association of four HC tropocollagens at low ionic strength. (A) Van der Waals representation of the systems with four HC tropocollagens after MD 200 ns simulations at low ionic strength. Tropocollagen fragments are indexed from T0 to T3, according to their positions in the rat tail collagen sequence, from N-terminal to C-terminal. Charged residues and hydroxyproline (Hyp) are highlighted with residue type coloring (see Fig. 3). (B) Contact map of the last structure (200 ns). (C) Interactions plots of first contacts in pairs T0-T1, T1-T2 and T2-T3; the different interaction types are colored as in Fig. 3.

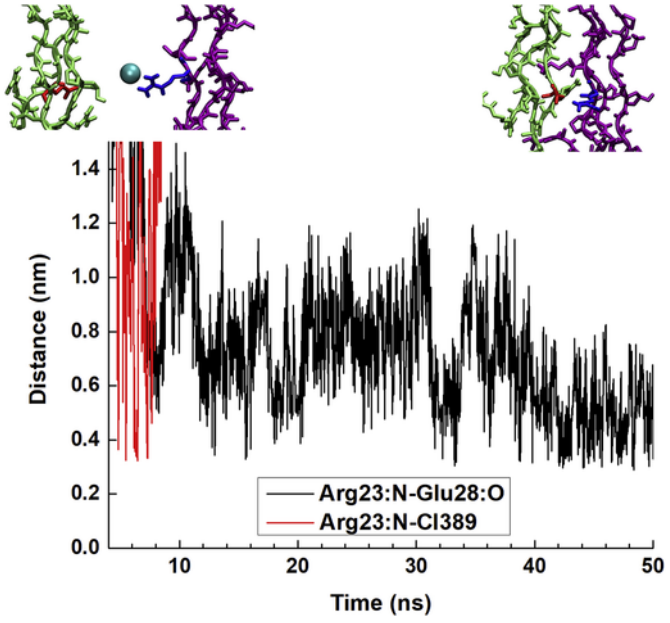

Fig. 14. Salt bridge formation after ion displacement at physiological ionic strength. Salt bridge formation between Arg23 of T1 fragment and Glu28 of T3 during the MD simulation with three HC tropocollagens at physiological ionic strength. The distance plot shows that the salt bridge can form only after the displacement of a Cl ion away from arginine. In the images on the top are two representative frames of this behavior; T1 fragment is in purple with Arg23 in blue and T3 fragment is in lime with Glu28 in red. The Cl ion (cyan) is in Van der Waals representation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

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

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

[3] G. Nemethy, H.A. Scheraga, Stabilization of collagen fibrils by hydroxyproline, Biochemistry (1986), doi:10.1021/bi00359a016.

[4] S. Perret, C. Merle, S. Bernocco, P. Berland, R. Garrone, D.J.S. Hulmes, M. Theisen, F. Ruggiero, 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.

[5] W. Pompe, H. Worch, W.J.E.M. Habraken, P. Simon, R. Kniep, H. Ehrlich, P. Paufler, Octacalcium phosphate-a metastable mineral phase controls the evolution of scaffold forming proteins, J. Mater. Chem. B (2015), doi:10.1039/c5tb00673b.

[6] M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu. Rev. Biochem. (2009), doi:10.1146/annurev.biochem.77.032207.120833.

[7] Y. Nomura, M. Yamano, C. Hayakawa, Y. Ishii, K. Shirai, Structural property and in vitro self-assembly of shark type I collagen, Biosci. Biotechnol. Biochem. (2009), doi:10.1271/bbb.61.1919.

[8] P. Noitup, M.T. Morrissey, 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.

[9] M. Thuanthong, N. Sirinupong, 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.

[10] J. Zhu, L.J. Kaufman, Collagen I self-assembly: revealing the developing structures that generate turbidity, Biophys. J. (2014), doi:10.1016/j.bpj.2014.03.011.

[11] E.E. Antoine, P.P. Vlachos, M.N. Rylander, Tunable collagen I hydrogels for engineered physiological tissue micro-environments, PLoS One (2015), doi:10.1371/journal.pone.0122500.

[12] I. Streeter, N.H. De Leeuw, A molecular dynamics study of the interprotein interactions in collagen fibrils, Soft Matter (2011), doi:10.1039/c0sm01192d.

[13] Z. Xu, Y. Yang, W. Zhao, Z. Wang, W.J. Landis, Q. Cui, N. Sahai, Molecular mechanisms for intrafibrillar collagen mineralization in skeletal tissues, Biomaterials (2015), doi:10.1016/j.biomaterials.2014.10.048.

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

[15] N. Krishnamoorthy, M.H. Yacoub, 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.

[16] S. Morozova, M. Muthukumar, Electrostatic effects in collagen fibril formation, J. Chem. Phys. (2018), doi:10.1063/1.5036526.

[17] EMBL, SIB Swiss Institute of Bioinformatics, Protein information resource (PIR), UniProt, Nucleic Acids Res, 2013, p. 41 (D43-D47).

[18] The Protein Data Bank, Struct. Bioinforma, 2005, doi:10.1002/0471721204.ch9.

[19] J.P.R.O. Orgel, T.C. Irving, A. Miller, T.J. Wess, Microfibrillar structure of type I collagen in situ, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 9001, doi:10.1073/pnas.0502718103.

[20] K. Kawahara, Y. Nishi, S. Nakamura, S. Uchiyama, Y. Nishiuchi, T. Nakazawa, T. Ohkubo, Y. Kobayashi, Effect of hydration on the stability of the collagen-like triple-helical structure of [4(R)-hydroxyprolyl-4(R)-hydroxyprolyl-glycine]10, Biochemistry (2005), doi:10.1021/bi051619m.

[21] J.A. Fallas, V. Gauba, 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.

[22] N. Guex, M.C. Peitsch, Swiss-model and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.

[23] C.D. Petrov, 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.

[24] C. Margreitter, M.M. Reif, 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.

[25] D. van der Spoel, E. Lindhal, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible and free, J. Comput. Chem. 26 (2005) 1701–1718.

[26] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. Van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, Eur. Biophys. J. (2011), doi:10.1007/s00249-011-0700-9.

[27] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[28] D. Mercadante, F. Gräter, C. Daday, CONAN: a tool to decode dynamical information from molecular interaction maps, Biophys. J. (2018), doi:10.1016/j.bpj.2018.01.033.

[29] 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.

[30] Y. Li, A. Asadi, M.R. Monroe, 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.