Disordered Filaments Mediate the Fibrillogenesis of Type I Collagen in Solution

Disordered Filaments Mediate the Fibrillogenesis of Type-I Collagen in Solution

Andrew R. McCluskey[1,2,§,‡], Kennes S. W. Hung[1‡], Bartosz Marzec[1†], Julien O. Sindt[2], Nico A. J. M. Sommerdijk[3], Philip J. Camp[1], and Fabio Nudelman[1]*

[1] EaStCHEM, School of Chemistry, The King's Buildings, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, UK

[2] EPCC, University of Edinburgh, Bayes Centre, 47 Potterrow, Edinburgh EH8 9BT, UK

[3] Department of Biochemistry, Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 6525 GA Nijmegen, The Netherlands

Keywords: cryo-electron microscopy, molecular simulations, fibrillogenesis, self-assembly, soft matter

ABSTRACT: Collagen type I is one of the major structural proteins in mammals, providing tissues such as cornea, tendon, bone, skin, and dentin with mechanical stability, strength, and toughness. Collagen fibrils are composed of collagen molecules arranged in a quarter-stagger array that gives rise to a periodicity of 67 nm along the fibril axis, with a 30-nm overlap zone and a 37-nm gap zone. The formation of such highly organized fibrils is a self-assembly process where electrostatic and hydrophobic interactions play a critical role in determining the staggering of the molecules with 67-nm periodicity. While collagen self-assembly has been extensively studied, not much is known about the mechanism, and in particular, the nature of the nuclei that initially form, the different stages of the aggregation process, and how the organization of the molecules into fibrils arises. By combining time-resolved cryo-transmission electron microscopy with molecular-dynamics simulations, we show that collagen assembly is a multi-step process in which the molecules first form filaments which self-organize into fibrils with a disordered structure. The appearance of the D-band periodicity is gradual and starts with the alignment of adjacent filaments at the N-terminal end of the molecules, first leading to bands with a periodicity of 67 nm and then to the formation of gap and overlap regions.

INTRODUCTION

Collagen type I is a major structural element in the extracellular matrix of tissues, such as skin, bone, dentin, and cornea, providing them with mechanical stability, strength, and toughness. The basic building block of the collagen fibril is the collagen molecule, which is a triple helix composed of two α1 chains and one α2 chain characterized by a repeating sequence of Gly-X-Y, where X and Y are often proline and hydroxyproline, respectively. Each molecule is approximately 300 nm in length and 1.5 nm in diameter. The molecules are organized into microfibrils, which build the collagen fibrils. Each microfibril is composed of 5 molecules arranged in a quarter-staggered array that gives rise to a D- period of 67 nm. This periodicity is due to the molecules being staggered with respect to each other in the axial direction by 67 nm, and results in a 30 nm overlap zone (0.46D) and a 37 nm gap zone (0.54D) [1-3]. This arrangement is evident at the fibrillar level when fibrils are visualized by transmission electron microscopy, where the overlap zones appear as dark bands, and the gap zones as light bands 2.

One of the most interesting features of collagen type I is the remarkable ability of the molecules to form such highly organized fibrils through self-assembly. The thermodynamic driving force for this process has a large entropic component, arising from the loss of surface-bound structured water. At low concentrations assembly follows a nucleation and growth mechanism with the extent of self-assembly following a typical sigmoidal curve [4-7]. Nuclei form in the lag phase, and then grow through axial and lateral association during the exponential phase. At high concentrations (above 100 mg/ml), assembly occurs through the formation of a cholesteric liquid crystalline phase, where both the concentration of the molecules and the viscoelasticity of the medium play a role in regulating fibril diameter[8, 9].

The organization of the collagen molecules into structures with a 67-nm D-periodicity is due to the staggering of the triple helices by multiples of 234 residues[10, 11], resulting in the maximization of the electrostatic and hydrophobic interactions between the molecules. Furthermore, the telopeptides – non-helical peptides at the N- and C-termini of the triple helices – also play a critical role in the self-assembly. Their proteolytic cleavage or the use of inhibitors significantly increases the time required for fibrillogenesis, inhibiting the axial and lateral growth of the fibrils, and the ones that eventually form have no distinguishable D-bands, meaning that there is no order in the packing of the microfibrils within the structure [4, 12-15].

The mechanisms of collagen self-assembly have been extensively studied over the years. It is known that the self-assembly is dependent on collagen concentration, pH, ionic strength, and temperature of the solution[16-19], and that these factors determine the morphologies and organization of the collagen fibril networks, microfibril density, D-periodicity, and collagen-collagen and collagen-substrate interactions [20-24]. One of the challenges in understanding collagen self-assembly, though, has been to identify the different stages and the intermediates in this process. Most studies on collagen self-assembly were performed at low concentrations (less than 10 mg/ml), where collagen has been proposed to assemble in three steps: nucleation, axial growth, and lateral growth[25]. The nucleation phase would involve the formation of intermediates, proposed to consist of linear quarter- staggered units varying from dimers and trimers to structures containing between 5 and 17 molecules packed laterally[26-28]. Their axial growth to form filaments 5-10 nm thick would follow, and their subsequent lateral association would lead to mature fibrils. More recent studies have suggested that microfibrils constitute the building blocks of collagen fibrils, with the formation of the 67-nm D-band apparent at very early stages[29, 30].

Despite all of the advances in understanding the factors that govern collagen self-assembly and the different stages in the process, the detailed mechanism is still largely unknown. There is still little information on the nature of the nuclei that initially form during the lag phase, and it is not known how the nuclei and filaments self-organize into ordered fibrils with the characteristic D-periodicity. Understanding the mechanisms of collagen self-assembly is of crucial importance, given that its structure and the proper alignment of the molecules in the quarter-staggered arrangement are critical for the integrity of several tissues, and for its ability to template hydroxyapatite during osteogenesis. Here, we used cryo-transmission electron microscopy (cryoTEM) to perform a time-resolved study on the mechanisms of self-assembly of collagen molecules into fibrils in solution. In combination with coarse-grained molecular-dynamics (MD) simulations, we show that collagen self-assembly starts with the formation of filaments that are 5 molecules-thick, and devoid of long-range order. These filaments then associate to form disordered fibrils, which eventually evolve into ordered structures with the typical D-band periodicity. Therefore, an important aspect of this work is the use of complementary experimental and simulation techniques to build up a complete picture of collagen fibrillogenesis.

MATERIALS AND METHODS

Collagen assembly: Acid-soluble bovine telopeptide-containing collagen solution (TeloCol, Advanced Biomatrix) or bovine collagen devoid of telopeptides (PureCol, Advanced Biomatrix) was diluted into 50 mM Tris buffer pH 7.4 containing 150 mM of NaCl to a concentration of 200 µg/ml and incubated at 37 °C. Samples were either incubated in a water bath and aliquots were

collected at different time points and frozen in liquid ethan for cryoTEM analysis, or placed in a UV-vis spectrophotometer (see below).

Staining of collagen with sodium phosphotungstate: Staining of collagen with sodium phosphotungstate was performed immediately before vitrification. Aliquots of 50 µl were collected from the reaction, mixed 1:1 in 2 % sodium phosphotungstate (Sigma-Aldrich), incubated for 15 seconds and plunge-frozen in liquid ethane.

Turbidity measurements: The turbidity measurements were performed using a VARIAN Cary 50 UV-vis spectrophotometer with the absorbance at 310 nm at 37 °C being measured every 2 minutes for the duration of the experiment.

CryoTEM analysis: R2/2 Quantifoil cryoTEM Au grids, grids (Quantifoil Micro Tools GmbH) were surface plasma treated for 45 seconds using a Quorumtech Glow Discharge system or a Cressington 208 carbon coater prior to use. Aliquots of 3 µl were taken from the reaction solution, applied to a cryoTEM grid and plunge-frozen in liquid ethane using a vitrification robot (FEI Vitrobot Mark III and IV), at 21 °C and 100 % humidity. Imaging was done using an FEI F20 Tecnai transmission electron microscope equipped with a field-emission gun operating at 200 keV and a Gatan cryoholder operating at -170 °C was used. Images were recorded using an 8k x 8k CMOS TVIPS F816 camera. Alternatively, a TU/e CryoTitan transmission electron microscope (FEI) equipped with a field emission gun operating at 300 keV, a post-column Gatan Energy Filter (GIF) and a post-GIF 2k x 2k CCD camera was used.

Image analysis: Mass-density profiles were measured using the line profile feature in Gatan Digital Micrograph. Measurements were taken along the length of a fibril and each point was integrated across the width of the fibril.

MD simulations: A coarse-grained collagen molecule was constructed from 200 beads with diameter $\sigma = 1.5$ nm (overall length 300.0 nm) and divided in to nine alternating segments (see Fig. 2a) of 20 beads (30.0 nm) and 25 beads (37.5 nm) with bond length 1.5 nm. The interactions between beads were expressed in terms of the Lennard-Jones (LJ) potential $u_{\mathrm{LJ}}(r) = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}]$ where r is the distance between beads, and ε and σ are the energy and range parameters, respectively. Attractive interactions, given by the cut-and-shifted potential $u_{\mathrm{att}}(r) = u_{\mathrm{LJ}}(r) - u_{\mathrm{LJ}}(2.5\sigma)$ for $r \leq 2.5\sigma$, operate between beads in the following pairs of segments: 1 and 3; 2 and 4; 3 and 5; 4 and 6; 5 and 7; 6 and 8; 7 and 9; and 9 and 1. All other interactions are repulsive, and are given by the Weeks-Chandler-Andersen (WCA) potential $u_{\mathrm{rep}}(r) = u_{\mathrm{LJ}}(r) - u_{\mathrm{LJ}}\left(2^{1/6}\sigma\right)$ for $r \leq 2^{1/6}\sigma$. The stiffness of the molecule is controlled with harmonic bond-stretching and bond-bending potentials with a spring constant of 2000ε in both cases. $N = 500$ molecules (with 10^5 beads) were placed randomly in a cubic box with length $L = 792$ nm, corresponding to a concentration of 1.67 μM; for a molecular mass of 300 kDa, this equates to $500\ \mu\mathrm{g\ ml^{-1}}$, a similar value to that used in the experiments. Periodic boundary conditions were applied. MD simulations were carried out with the solvent being represented implicitly by random Brownian forces and viscous (Stokes) forces acting on the beads. The calculations used reduced units: the dimensionless temperature is $T^* = k_{\mathrm{B}}T/\varepsilon$, where k_{B} is Boltzmann's constant; the reduced concentration is $c^* = N\sigma^3/V = 3.40 \times 10^{-6}$, where $V = L^3$ is the box volume; and the reduced time is $t^* = t\sqrt{\varepsilon/m\sigma^2}$ where m is the nominal bead mass.

Simulations were carried out at reduced temperature $T^* = 1$, with Stokes friction coefficient $\gamma^* = 1$, and with reduced time step $\delta t^* = 0.01$. A run of 0.2×10^9 time steps ($t^* = 0.2 \times 10^7$) with repulsive interactions only was used to equilibrate a disordered structure, and then 1.9×10^9 time steps ($t^* = 1.9 \times 10^7$) were performed with the attractive interactions defined above to simulate the self-assembly process. All simulations were carried out using LAMMPS[31,32].

RESULTS

Time-resolved study of collagen self-assembly: The self-assembly of collagen molecules into fibrils was induced by diluting type-I collagen dissolved in 10 mM HCl at pH 2 with a Tris-buffered saline solution (150 mM Tris buffer and 150 mM NaCl) to a final concentration of $200\ \mu g\ ml^{-1}$ at pH 7.4 and 37 °C. This concentration was chosen so that we could resolve the growth and development of individual fibrils using cryoTEM during the nucleation and growth stages. Two forms of commercially available type I collage were used: TeloCol (Advanced Biomatrix), which contains the telopeptides, and PureCol (Advanced Biomatrix), which had the telopeptides removed. The fibrillogenesis process was monitored by measuring the increase in the turbidity of the solution by UV-vis spectrophotometry at 310 nm. As previously reported for type-I collagen 5, the turbidity curves for the telopeptide-containing collagen followed a typical sigmoidal pattern, with three stages: a lag-phase period with no detectable increase in turbidity, lasting approximately 7 minutes; a growth phase in which the turbidity increased rapidly; and a plateau region where turbidity did not change any more (Fig. 1a). No increase in turbidity was observed for the collagen devoid of telopeptides (Fig. 1a), indicating that no self-assembly took

place. Therefore, subsequent cryoTEM measurements were done only on telopeptide-containing collagen.

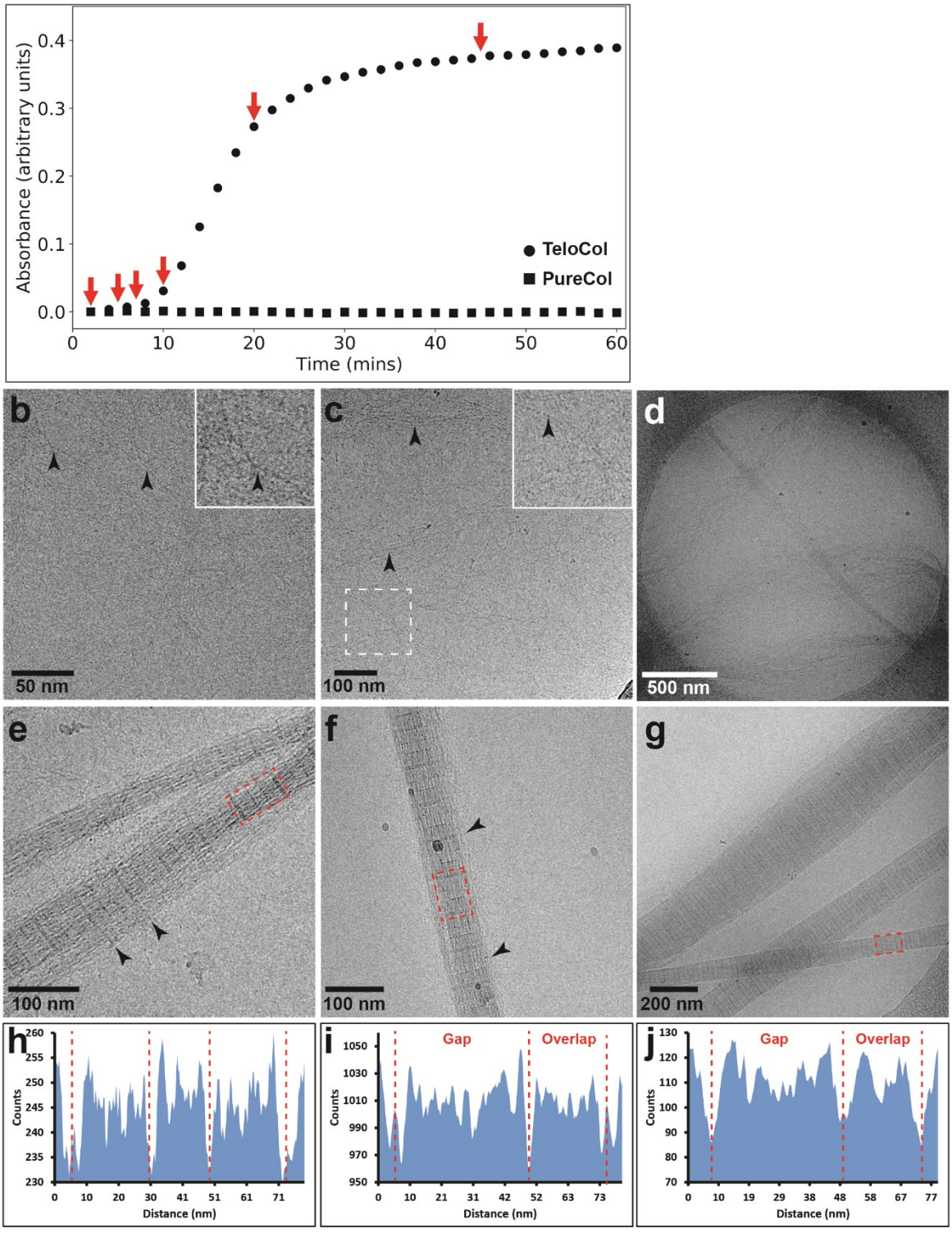

Figure 1. Turbidity curve of TeloCol (telopeptide-containing-collagen, black circles) and PureCol (atelo-collagen, black squares) (a) and time-resolved cryo-transmission electron microscopy (cryoTEM) images of the different stages of TeloCol self-assembly (b-g). a) Increase in turbidity over time during TeloCol assembly in Tris-buffered saline solution at 37 °C. Red arrows indicate time points when samples were collected for cryo-transmission electron microscopy analysis. b) After 5 minutes of reaction line-like structures 2 nm-thick were present (black arrowheads). Inset: higher magnification of the area marked by the dashed square. c) Image taken after 7 minutes, where filaments 4 nm-thick were starting to organise into fibrillary structures (black arrowheads). Inset: higher magnification of the area marked by the dashed square. d) After 10 minutes, where disordered, loosely packed fibril-like structures were present. e) Collagen fibrils imaged after 15 minutes. The D-banding of collagen started to develop (dotted square) at the core of the fibril. Black arrowheads: bands extending from the well-organised core to the poorly organised periphery. f) Fibril imaged after 20 minutes had well-defined gap and overlap regions (dotted square). Black arrowheads: poorly ordered region at the periphery of the fibril, containing 4 nm-thick filaments. g) Mature fibrils present after 40 minutes of reaction. h), i) and i) Mass-density profiles of the areas marked by the dotted squares in e), f) and g), respectively. Valleys in the mass-density profile (lower counts) correspond to darker areas in the cryoTEM images, such as dark bands separating between the different regions in a 67-nm repeat.

Using the turbidity curve as a guide, samples were collected at different time points during the lag, growth, and plateau phases (Fig. 1a, arrows), plunge-frozen in liquid ethane, and analyzed by

cryoTEM. This allows the imaging of collagen during the different stages of self-assembly, ensuring the preservation of the molecular structures in their close-to-native, hydrated states[33].

After 2 minutes of reaction, the collagen was still present mostly in its molecular form, as no features or objects were visible in the cryoTEM images (see Supporting Information 1, Fig. S1). Individual collagen molecules, about 1.5 nm thick, do not provide enough contrast to be visible within the amorphous ice layer under the imaging conditions used. However, after 5 minutes, aggregates were present in the solution, evident by the observation of line-like structures with a thickness of 2.2 ± 0.5 nm (Fig. 1b, arrowheads). These structures were significantly thinner and shorter than the 5-10 nm filaments previously imaged by conventional TEM 20, 25, and it is conceivable that they constitute collagen dimers and trimers, as proposed before [27, 34]. The onset of fibril formation corresponded to the start of the exponential phase in the turbidity curve, at 7 mins. At this stage, the molecular aggregates had further self-assembled into a network of filamentous structures 4.0 ± 0.6 nm in diameter (Fig. 1c, arrowheads) and hundreds of nanometers long that were aligned into parallel arrays. Although these filaments had the same thickness as a microfibril, suggesting that they are bundles of 5 molecules, we do not know if they had the same quarter-staggered organization. After 10 minutes of reaction these filaments developed into disorganized, loosely packed pre-fibrillar structures spanning a few microns in length (Fig. 1d). These pre-fibrillar structures can be seen to further self-organize into compact fibrils 30 to 60 nm thick, concomitant with the appearance of the characteristic banding pattern of collagen. Indeed, Fig. 1e depicts a fibril that has two distinct regions: a well-organized core with an average D-band spacing of 68 nm (Fig. 1e, red box), and a poorly organized periphery. At this stage, it is difficult to distinguish between the gap and overlap zones within the 68-nm banding. Instead, these repeats are divided into three bands of similar contrast, with two bands measuring 25 nm separated by a

lighter band 18 nm wide (Fig. 1e and 1h). While the periphery of the fibril did not yet exhibit a banding pattern, bands extending from the well-ordered core could already be distinguished (Fig. 1e, arrowheads). Additionally, it can be seen that the forming fibrils are composed of filaments approximately 4 nm thick, similar to the ones found at 7 minutes (see Supporting Information 2, Fig. S2).

Fibrils observed after 20 minutes of reaction were between 50 and 100 nm in width and were compact and well organized showing a well-defined banding pattern with distinct gap and overlap zones (Figs. 1f and 1i). Interestingly, the length of the D-banding at this stage was longer than at 10 minutes: 71 nm, with the gap region measuring 45 nm, and the overlap region measuring 26 nm. In addition, filaments 4 nm-thick were also present, merging with the fibril at its periphery and forming a thin, disordered layer (Fig. 1f, arrowheads). Averaging the mass-density profiles of the fibrils present after 15 and 20 minutes to better characterize their banding patterns[35, 36] was not possible since at this stage of self-assembly they were still forming and not all identical.

After 45 minutes, which corresponded to the plateau region in the turbidity curve, the fibrils were fully mature, measuring 100–300 nm in diameter. Moreover, the D-spacing now measured 67 nm, with a gap zone of 39 nm and an overlap zone of 28 nm (Figs. 1g and 1j). These values are in the range of those found in previous reports[1, 3].

Our data show two self-assembly processes occurring during collagen fibrillogenesis. The first stage is the clustering and organization of the molecules into 4 nm-thick filaments. We suggest that these filaments are 5 molecules-thick, similar to microfibrils that constitute the basic unit of collagen fibrils. However, we cannot determine if within these filaments the molecules are organized in the quarter-stagger arrangement or not, and hence whether they already have the microfibril structure 1 or not. In the second stage, these filaments align and self-organize into

disordered pre-fibrillar structures, which gradually anneal into mature, ordered fibrils. To gain more insight on the mechanisms of this self-assembly process, both stages were further investigated with two complementary approaches covering the different size and time domains. As the first stage is not directly observable using microscopy techniques, MD simulations of a simplified model were used to gain insights on the very first steps of collagen fibrillogenesis. The second stage involves the formation of fibrils which are directly observable using cryoTEM methods, but the timescales (minutes) are inaccessible to molecular-level simulations.

Stage 1 – filament formation: To make a relevant model for the assembly of collagen we should consider the following: 500 collagen molecules contain about 21 million atoms. At the experimental concentration, the cubic box length is L = 792 nm, which contains about 17 billion water molecules. Therefore, to simulate the self-assembly of collagen with atomic or chemical resolution over the relevant length and time scales is impossible. In coarse-grained models, only the most relevant molecular details are retained in order to describe the phenomena of interest. In the present case, this necessarily means that chemical details, atomic-scale structure, solvent, etc. are omitted from the model. This type of strategy has been followed to study collagen assembly on surfaces[24] as well as in the biophysical modelling of several other processes[37] such as genome packing[38], protein folding[39], protein aggregation[40], DNA melting[41], DNA unzipping[42], amyloid aggregation[43], virus capsid self-assembly[44], and biomineralization[45].

A minimal coarse-grained model was therefore constructed to give the polar ordering of collagen molecules and with the correct quarter-staggering between pairs of molecules (Fig. 2). Each molecule was represented by a linear chain of 200 beads with diameter $\sigma = 1.5$ nm, divided up into segments to reflect the lengths, diameters, overlap zone, and gap zone of collagen

molecules (Fig. 2a). The interactions between beads in particular segments were either attractive or repulsive so as to stabilize the correct ordering of molecules within clusters. Nine segments were used so that the energy of parallel (polar) ordering of the molecules was much lower (more favorable) than that of anti-parallel ordering. It is important to note that in this coarse-grained model the segments were not intended to represent the real amino-acid sequence, charge distribution, or van der Waals interactions. However, to do anything else is currently not feasible. The solvent was represented implicitly as a continuum. MD simulations of $N = 500$ collagen molecules at the experimental concentration were carried out under constant-temperature conditions using the LAMMPS software package[31, 32]. The self-assembly of the molecules in a viscous liquid under isothermal conditions was thereby simulated, corresponding to the experimental conditions. If the interaction model contains the essential features of the real molecules, then the simulated mechanism should give relevant insights on the self-assembly pathway.

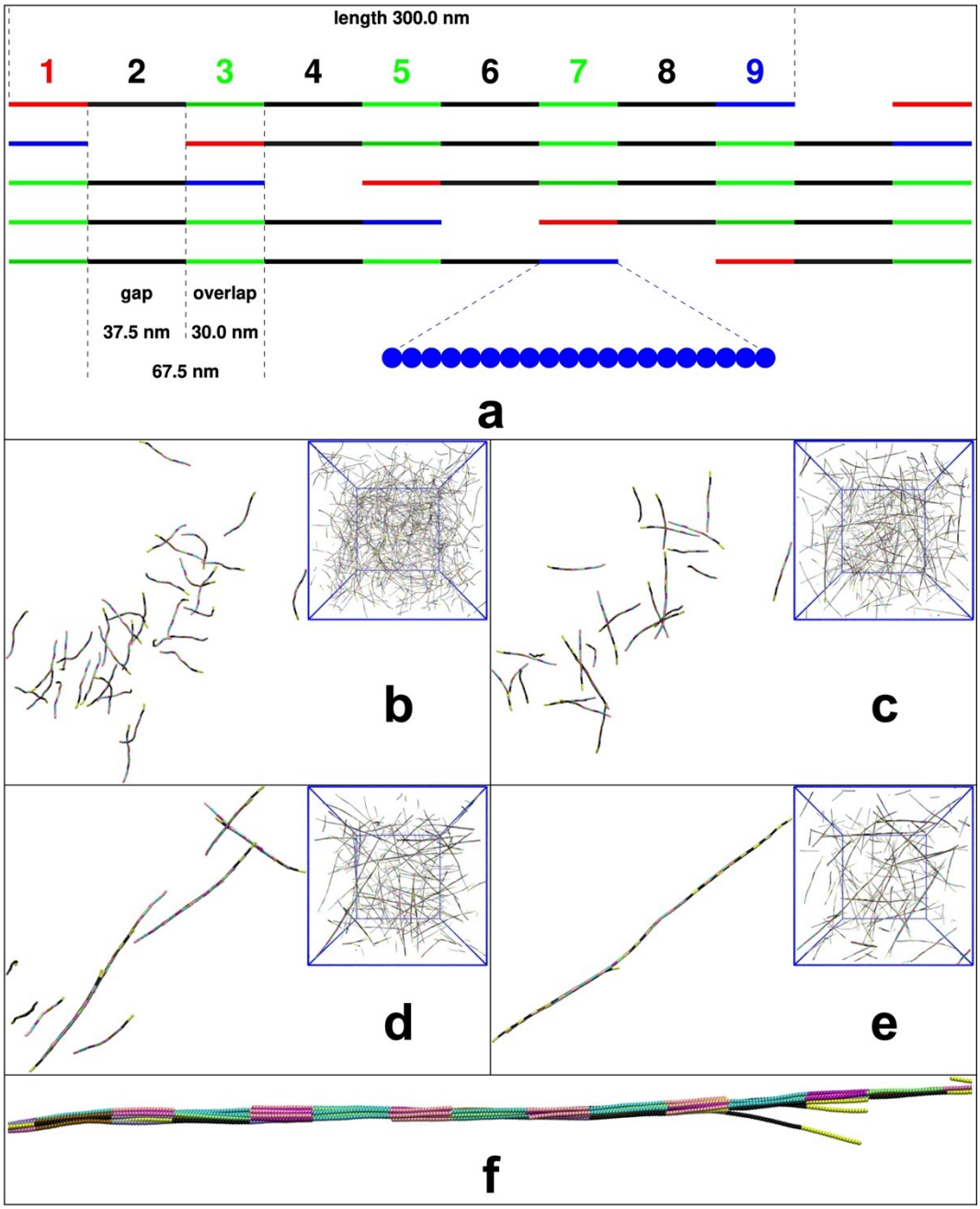

Figure 2. The coarse-grained molecular model used in simulations. a) Schematic diagram of coarse-grained collagen molecules showing segments 1-9 aligned in a low-energy configuration. Each segment is made up of a chain of beads interacting via attractive or repulsive potentials. The interaction between two coloured (30.0 nm) segments (1,3,5,7,9) or two black (37.5 nm) segments (2,4,6,8) is attractive, while the interaction between a black segment and a coloured segment is

repulsive. The red and blue segments indicate the polar ordering of the molecules, and the other short segments are shown in green. b)–e) Snapshots of the system at different simulation times: b) $t^* = 0.0$; c) $t^* = 0.4 \times 10^7$; d) $t^* = 0.9 \times 10^7$; e) $t^* = 1.9 \times 10^7$. The main panels show only those molecules which form the largest cluster at the end of the simulation ($t^* = 1.9 \times 10^7$). Insets show the whole system at low magnification; the box length is $L = 792$ nm. f) High magnification of the largest cluster formed at $t^* = 1.9 \times 10^7$.

Despite its simplicity, the model results in staggering between molecules where the overlap and gap distances are well in line with the experimental values for collagen [3, 36]. Indeed, while at the beginning of the simulation (Fig. 2b) the molecules were completely dissociated from each other, during the simulation the combined interactions within the model give rise to dimers with parallel (and polar) ordering of the molecules and their axial staggering by 67 nm, with overlap and gap distances of 30.0 nm and 37.5 nm, respectively. Subsequently the simulation showed the formation of, trimers, tetramers etc., and ultimately filaments (Figs. 2c–2e and supplementary movies 1 and 2), while the proportion of dissociated monomers decreased. Importantly, filament growth occurred predominantly in the axial direction, while growth in the lateral direction was limited.

The assembly of the collagen molecules was monitored through the evolution of the radial distribution function (RDF) $g(r)$ of the centers of mass (COMs) of the segments (Fig. 3a). The RDF measures the probability of finding two segment COMs separated by a distance r [46], and reflects the progress of fibrillogenesis as the segment COMs come into registry during collagen self-assembly. At the start of the simulation, the first two peaks were at $r = 33.7$ nm and $r = 67.5$ nm, which correspond to the distances between segment COMs on the same molecule: the first peak is at the average of the short-segment and long-segment lengths, meaning adjacent segments; and the second peak is at the sum of these two lengths, meaning two segments separated

by a third. For all other distances in the range $0\ \text{nm} \leq r \leq 75\ \text{nm}$, it was seen that $g(r) = 1$, indicating that there were no correlations between segment COMs on different molecules. Hence, there was a random distribution of the collagen molecules at the start of the simulation.

At the end of the simulation, there was an intense and broad band at $r \leq 15\ \text{nm}$, which signaled strong intermolecular association, since there is no other way that segment COMs can get that close to one another; recall that the COMs of adjacent segments on the same molecule are separated by $33.7\ \text{nm}$. The peaks near $33.7\ \text{nm}$ and $67.5\ \text{nm}$ were broadened due to a thermal distribution of distances between segment COMs on neighboring molecules, resulting from molecules being a few bead diameters away from the perfect, lowest-energy alignment; the intramolecular contributions to these peaks were, of course, unaffected by such thermal fluctuations.

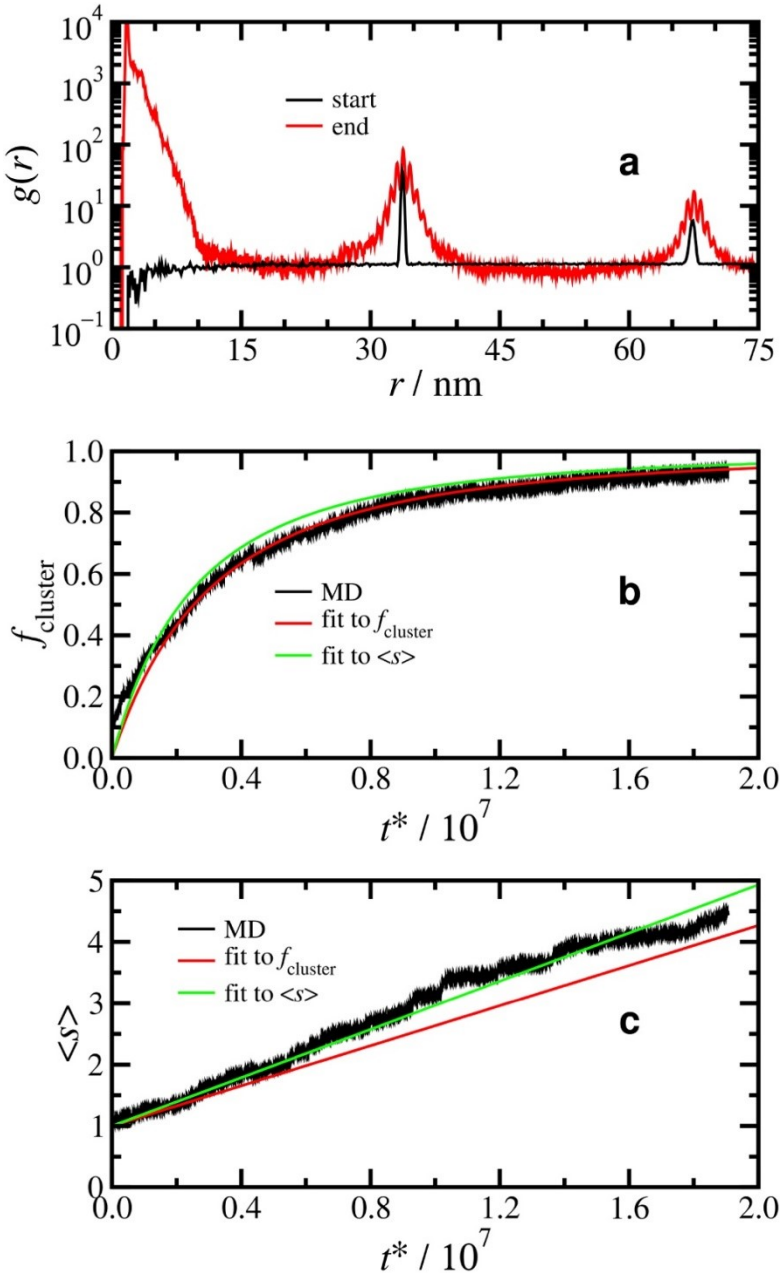

Figure 3. Results from MD simulations. a) Radial distribution functions [RDF or $g(r)$] of the segment centers of mass at the start of the simulation (black) and at the end of the simulation (red). The peak at 33.7 nm represents the average of the short-segment and long-segment lengths, and the peak at 67.5 nm represents the sum of the two segment lengths, both reflecting the intramolecular structure. b) The fraction of particles in clusters, f_{cluster}, as a function of time from simulation (black) and as fitted by the von Smoluchowski model (red and green). c) The average

cluster size $\langle s \rangle$ from simulation (black) and as fitted by the von Smoluchowski model (red and green). The extrapolation of the average cluster size to $t^* = 2.0 \times 10^7$ is $\langle s \rangle = 4.93$. In b) and c), the red curves are with a rate coefficient fitted to f_{cluster}, and the green curves are with a rate coefficient fitted to $\langle s \rangle$.

Given the very low concentration, there is negligible probability of two randomly oriented molecules having segment COMs in contact (Fig. 3a), which makes this a sensitive measure for assembly formation. Therefore, two molecules were considered clustered (in dimers or larger structures) if at least two segment COMs had a separation of less than 15 nm. During the simulations, the fraction of molecules in clusters, f_{cluster}, gradually increased with ~95% of the molecules being part of clusters by the end of the simulation ($t^* = 1.9 \times 10^7$, Fig. 3b). Applying simple von Smoluchowski kinetics of sequential cluster growth [47] (Supporting Information 3), we can derive a time-dependent average cluster size $\langle s \rangle$ (Fig. 3c). Extrapolation of $\langle s \rangle$ gives an average cluster size $\langle s \rangle = 4.93$ at $t^* = 2.0 \times 10^7$, while the approximately constant slope of the plot time suggests that further growth of the clusters will occur, either through further association of the existing clusters, or by a dissociation-reassociation mechanism similar to an Ostwald ripening process. At $t^* = 1.9 \times 10^7$ a wide range of cluster sizes was apparent. The largest filament contained approximately 30 molecules (270 segments), had a length of about 50 segments, and hence an average cross section of 5-6 molecules, although there were thick and thin sections (Fig. 2f). Assuming an approximately close-packed lateral arrangement of the molecules (Fig. 2f), the average diameter was ~4.5 nm. Although the molecules were designed to show a staggering between pairs well in line with the established dimensions of the gap and overlap regions in collagen, and the resulting diameters of the filaments were in good agreement with the

experimental observations, the simulations did not show evidence for long-range positional order within the filaments (Fig. 2f). Specifically, although the segments on different molecules have self-organized into bands, stabilized by the attractive interactions between segments defined in the coarse-grained model, the molecules do not exhibit any lateral or longitudinal order within filaments.

These simulations support the presence of a lag time during the early stages of fibrillogenesis, where turbidity measurements cannot detect any structure formation, but during which individual collagen molecules assemble into staggered dimers, trimers, etc. with diameters of 3.0 nm and above. Further axial and lateral combination of single molecules, dimers, trimers, etc. leads to the formation of 4.5 nm thick disordered filaments. These are distinct from the well-defined microfibrils found inside collagen fibrils, in which molecules are ordered into 'unit cells' with a quarter-staggered arrangement 1. Following this scenario, therefore, the quarter-stagger ordering of the collagen molecules must occur at a later stage, either during or after the assembly into the larger fibrils.

Stage 2 – from filament to fibril: In order to gain a more detailed insight into the mechanisms of assembly of the initial filaments into mature fibrils, we combined cryoTEM with sodium phosphotungstate staining. Sodium phosphotungstate binds to the positively charged amino acids of collagen, increasing the local mass density in those regions and allowing us to observe 11 distinct bands within one D-repeat (a through e) [48, 49]. The bands c2, c1,b2, b1, and a4 are located in the overlap region, and bands a3, a2, a1, e2, e1, and d are in the gap region. These bands have been correlated to the corresponding spatially close groups of positively charged amino acids in the amino-acid sequence of collagen [48, 49]. Staining collagen at different stages of formation

therefore provides detailed information on how during fibrillogenesis, the collagen molecules organize into a highly-ordered collagen fibril (Fig. 4).

Samples collected after 7 minutes and stained with sodium phosphotungstate showed the presence of a disorganized network of filaments 4.6 ± 0.7 nm in diameter (Fig. 4a). These were similar in size to the ones observed in unstained samples, but with significantly higher contrast due to the positive staining. After 10 minutes these filaments started to align into parallel arrays, leading to loosely packed pre-fibrillar structures. Although these fibrillar structures were up to 1 μm in width, their contrast in cryoTEM suggests that they were only a few fibrils thick, hence possibly residing at the air-water interface of the thin film prior to plunge freeze vitrification (Fig. 4b).

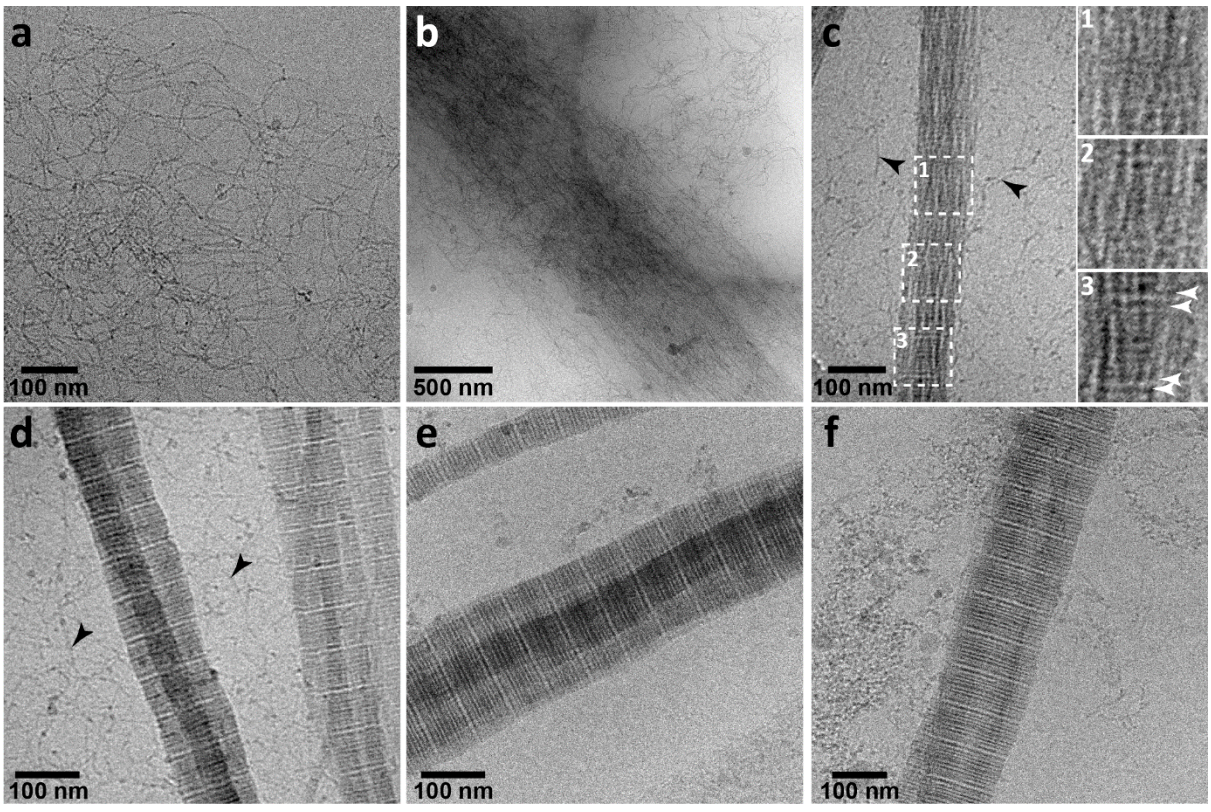

Figure 4. Time-resolved cryo-transmission electron microscopy (cryoTEM) images of the different stages of collagen self-assembly stained with sodium phosphotungstate. a) After 7

minutes of reaction a network of filaments 4 nm-thick was present. b) Image taken after 10 minutes, showing filaments aligned into parallel arrays, forming loosely packed fibril-like structure. c) Collagen fibril imaged after 15 minutes, with 4 nm-thick filaments in the surrounding medium (black arrowheads). Gradual stages of assembly can be seen in the fibril (insets 1, 2 and 3). White arrowheads in inset 3: doublets of staining bands. d) Collagen fibril imaged after 15 minutes at a more developed stage of assembly, where staining bands are more discernible. Black arrowheads: 4 nm-thick filaments in the surrounding medium. e) Fibril imaged after 20 minutes. f) Mature fibrils present after 40 minutes of reaction.

As with the unstained samples, after 15 minutes of assembly the fibrils were significantly more compact, with diameters of ca. 90 nm (Fig. 4c and Supporting Information 4, Fig. S3). Filaments of 4 nm in diameter were present in the surrounding medium, becoming incorporated into the fibrils and extending their diameter (Fig. 4c, black arrowheads and Supporting Information 4, Fig. S3). From this stage onwards, the positive staining revealed the progression of the ordering within the fibrils through the development of the staining bands. This is exemplified by Fig. 4c which shows a fibril containing three regions with distinct stages of organization. Region 1 displayed no visible organization and hence no staining bands could be distinguished (Fig. 4c, inset 1 and Fig. 5a). The first signs of the alignment of the microfibrils within the fibrils were visible in region 2, where two doublets of dark bands interspaced by a distance of 66 nm had appeared (Fig. 4c, inset 2) which were clearly visible in the corresponding mass-density profile (Fig. 5b, black arrowheads).

Importantly, this doublet - always separated by a distance of 66 nm - was a constant feature in the development towards the more ordered fibrils (Fig. 4c, inset 3, white arrowheads, and Fig. 5c,

black arrowheads). This shows, therefore, the first appearance of a repetitive structure corresponding to the D-staggering of the molecules within the fibril. Additionally, a number of staining bands, still ill-defined and within a strong background, were starting to emerge, indicating the development of the microfibrils within the (pre-)fibrillar structure. However, no gap and overlap regions could be distinguished at this stage.

Fibrils in a more advanced stage of assembly were also present after 15 minutes (Fig. 4d and Supporting Information 5, Fig. S4). They displayed a regular staining pattern with a D-band periodicity of ~ 66 nm, showing that the now collagen microfibrils had become aligned with respect to each other (Fig. 4d). This pattern translated into well-defined peaks in the mass-density profile (Fig. 5d) that could already be assigned to their respective classifications (c2, c1, etc.) with the exception of peaks a2, a3 and a4, which were still merged into one band. From this, it becomes evident that the doublet of bands that were first to appear (Fig. 4c, insets 2 and 3 and Figs. 5b and c) are the d and c2 staining bands. Correlating the position of the staining bands with the amino-acid sequence of collagen reveals that this region corresponds to the N-terminal end of the collagen molecules, at the border between the gap and overlap regions [48]. It indicates that, once a disorganized bundle of filaments forms, they start their alignment at the N-terminal region of the molecules to generate the D-periodicity that is characteristic of collagen fibrils. This telopeptide was suggested to be important for placing the D-periods in register during lateral growth of the fibrils through the association of the microfibrils 13. Our observations are well in line with that.

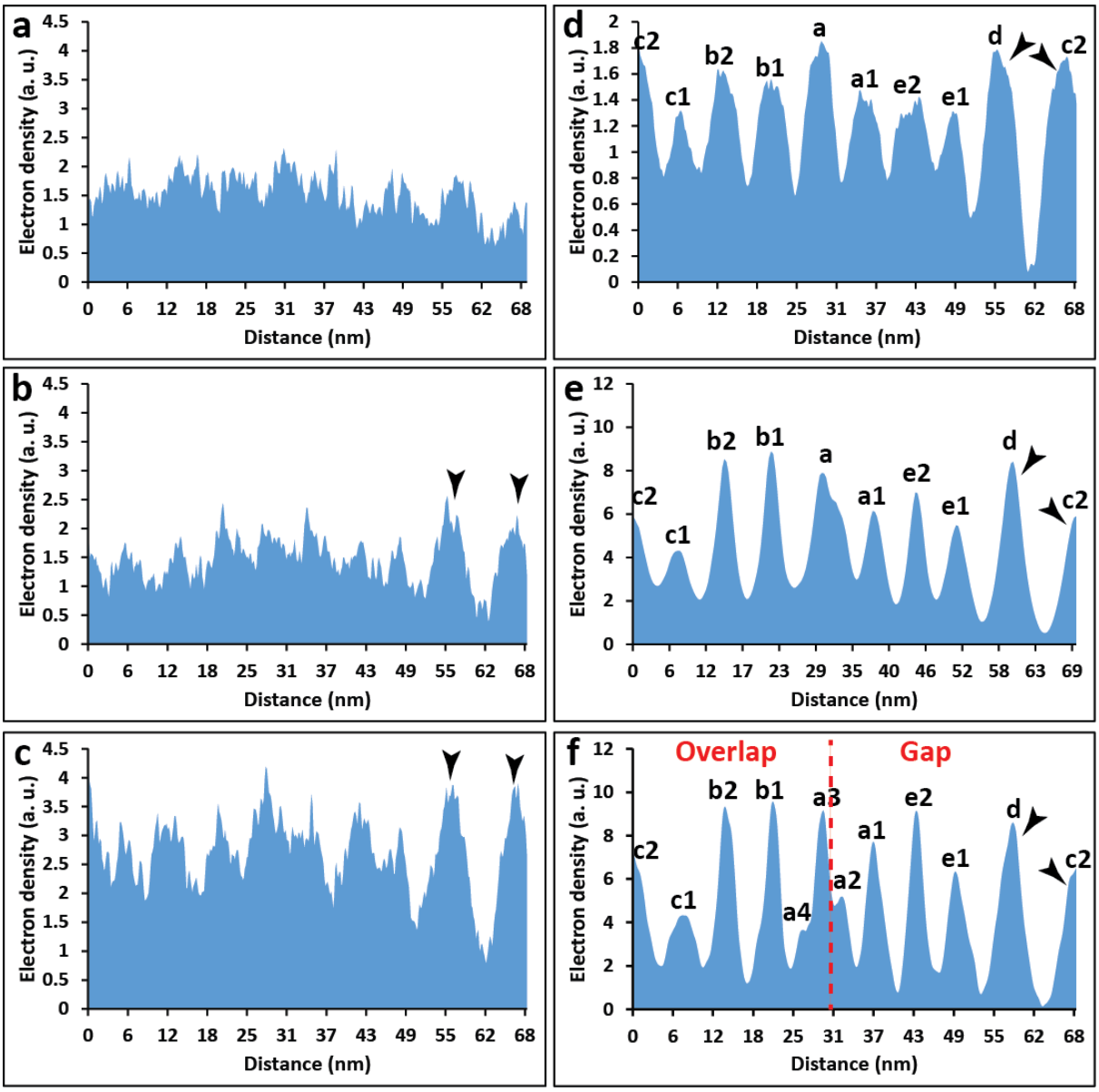

Figure 5. Mass-density profiles of the collagen fibrils stained with sodium phosphotungstate depicted in Fig. 5. Each peak corresponds to a dark staining band. a) Mass-density profile of inset 1 in Fig. 5c (15 minutes). b) Mass-density profile of inset 2 in Fig. 5c. c) Mass-density profile of inset 3 in Fig. 5c. d) Mass-density profile the collagen in Fig. 5d (15 minutes). e) Mass-density profile the collagen in Fig. 5e (20 minutes). f) Mass-density profile the collagen in Fig. 5f (40 minutes). Black arrowheads in b)-f): peaks corresponding to the doublet of staining bands

delimiting the D-banding. The peaks in d)-f) correspond to well-defined staining bands and were assigned to their classification of a-e2, according to [48].

After 20 min, the staining bands had become better defined, reflecting the increase in ordering and alignment of the adjacent microfibrils (Fig. 4e and Supporting Information 5, Fig. S4). Concomitantly the peaks in the mass-density profile, corresponding to the staining bands, became sharper and better defined (Fig. 5e). Additionally, at this point filaments were no longer present in the surrounding medium. After 40 minutes the fibril had become fully mature, with all characteristic staining bands present and being well-defined, demonstrating that the microfibrils were now well aligned within the fibrils (Figs. 4f and 5f).

As the assembly progressed, we analyzed the length of the D-banding repeat by fitting a sum of Gaussian functions to the mass-density profile of the different fibrils (Supporting Information 6, Fig. S5). After 15 minutes of assembly (Fig. 4d) this yielded a D-band periodicity of 65.9 ± 3.4 nm with an average full-width half-maximum (FWHM) of 4.6 ± 0.5 nm. Interestingly, after 20 minutes the length of the D-banding had increased to 70.2 ± 3.0 nm and an average FWHM of 4.1 ± 0.2 nm, consistent with the unstained collagen imaged at the same stage (Figs. 1e and 1h). Then after 40 minutes, the mass-density fits gave again a shorter D-band periodicity of 67.8 ± 2.7 nm with an average FWHM of 3.5 ± 0.3 nm. The decreasing average FWHM show the alignment of the microfibrils to form a highly ordered fibril. While the differences in the D-band periodicity, obtained from the fitting of a Gaussian function to the mass-density profiles, might not be statistically significant, they indicate variations in the length of the D-periodicity as a function of the stage of assembly.

DISCUSSION

We show that the self-assembly of type I collagen molecules into ordered fibrils is a multi-step process involving the evolution of long-range order. It starts with the assembly of collagen molecules into dimers/trimers and then filaments that are 4 nm thick and hundreds of nanometers long, which then align into uniaxially oriented, loosely-packed prefibrillar structures. This is consistent with the prevailing picture of formation of dimers/trimers, axial elongation, and then lateral thickening. The packing density of these bundles then increases, leading to the formation of fibrils still devoid of the 67-nm banding. The last stage is the appearance of the banding pattern within the fibrils. This is illustrated schematically in our proposed model in Fig. 6 as a structural evolution from molecules, through clusters, disordered filaments, and disordered fibrils, to ordered fibrils. Overall, this process is analogous to crystallization, where even the critical nuclei show some disorder [50] and amorphous phases precede a crystal lattice [51, 52]. Here, the disordered, loosely packed fibrils are comparable to an amorphous intermediate, and the emergence of long-range order within fibrils is analogous to the ordering of the ions or molecules in a crystal lattice. In the case of collagen, however, the ordering occurs via the rearrangement of the pre-fibrillar structures, rather than through dissolution and re-precipitation. The three main stages of collagen fibrillogenesis – the formation of 4 nm-thick filaments, their lateral aggregation into prefibrillar structures and their evolution into fibrils with a well-defined banding pattern – will be discussed separately below.

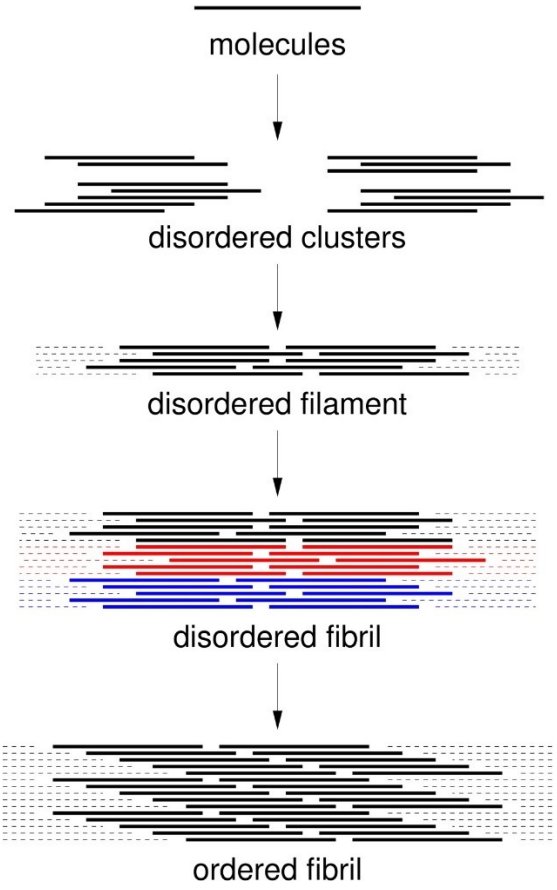

Figure 6. Schematic diagram of the self-assembly process.

Assembly of collagen molecules into filaments: Our cryoTEM observations show the presence of 2.2 ± 0.5 nm-thick line-like structures after 5 minutes of assembly. Given that the collagen molecules are ca. 1.5 nm in diameter and do not provide enough contrast to be visualized by cryoTEM individually, we propose that these structures correspond to collagen dimers and trimers. This interpretation correlates well with the MD simulations showing that collagen assembly starts with the aggregation of the molecules into staggered dimers and trimers. Probing the assembly further using MD simulations and cryoTEM, we show that these dimers and trimers aggregate laterally and longitudinally to form longer filaments that are around 5 molecules-thick and with a

diameter of 4 nm, similar in dimensions to microfibrils. Altogether, these observations are consistent with previous electron microscopy and dynamic light scattering measurements of collagen fibrillogenesis in vitro and in vivo [27, 34, 53].

An important aspect of the self-assembly of the collagen molecules into fibrils is the stage in which they develop long-range order, in other words, the 67-nm pattern that is characteristic of mature fibrils. Our MD simulations show that the filaments that form in the first stage of assembly are devoid of such long-range order. This means that while molecules may form staggered dimers and trimers, the further growth of these units does not seem to lead to a bundle of 5 or more molecules already containing the microfibrillar organization, as is evident from the simulation snapshots. While these data must be verified experimentally, the aggregation of molecules to form such thin, disordered filaments with high surface-to-volume ratios makes sense from the thermodynamic point of view.

Consider a dimer of collagen molecules in a low-energy, quarter-stagger arrangement. Where does a third molecule go? Provided that the strongest intermolecular interactions are short-ranged, the third molecule can occupy a number of low-energy configurations with respect to one of the other molecules, but not both; in the present case, if molecules 1 and 2, and molecules 2 and 3, have the ideal stagger, then molecules 1 and 3 will not have the ideal stagger. Hence, there is no energetic driving force for the third molecule to adopt the precise position that it would have in the microfibril. Because there is no single preferred cluster structure, a range of structures should be anticipated. To some extent, this argument can be extended to filaments provided that they are not too thick, meaning that the majority of molecules are not fully surrounded by neighbors as they would be in the bulk phase. Essentially, if the cluster of molecules is ‘small’, then the proportion

of molecules near the surface will be high, and because those molecules are not fully surrounded by other molecules, there will be many possible positions with comparable energies.

In a fibril, there should be a thermodynamic benefit to having all of the molecules in their respective 'unit-cell' positions. The cost of producing a defect, whereby one or more molecules are shifted from their preferred positions, would arise from interactions with all of the surrounding molecules. Hence, there should be a narrow and well-defined manifold of ground-state structures. The interactions of each molecule with its full complement of surrounding molecules could be the cause of the banding that develops after the initial formation of disordered fibrils. Indeed, Tresladt et al. observed the presence of a 67-nm repeat pattern only on fibrils that were at least ca. 20 nm in thickness, and even then they still contained regions devoid of a periodic staining pattern [54]. These results led to the conclusion that the initial fibrils were still loosely packed in order to allow readjustments of the molecules towards increased order.

The organization of disordered filaments into disordered fibrils, and the subsequent development of the banding pattern, are considered next.

Organization of filaments into fibrils: The next step in fibrillogenesis is the organization of the filaments into fibrils. We show that this process starts with the formation of uniaxially oriented bundles of loosely packed filaments, followed by their increase in packing density to generate compact fibrils. The mechanism for this initial alignment could be similar to that for the isotropic-nematic phase transition in highly elongated molecules, in which the transition is driven mainly by the entropic contributions to the free energy. In the seminal theory proposed by Onsager [55, 56], the loss of orientational entropy on alignment is compensated by the gain of translational entropy afforded by the reduction in the excluded volume of two molecules. This isotropic-nematic phase

transition brings individual filaments in close proximity, so that electrostatic and hydrophobic interactions between molecules from adjacent filaments can take place and mediate the formation of compact fibrils.

At this stage, these fibrils do not yet display the 67-nm periodicity, and based on our MD data we propose that the molecules are not yet in a quarter-stagger arrangement. These results are in contrast to AFM measurements that show the appearance of the D-banding of collagen at a very early stage, with the microfibrils already in register with the nascent fibrils when they fuse to the latter [30]. These differences in observations are possibly due to the different experimental conditions, in particular the presence of a mica surface that may accelerate fibrillogenesis and give directionality to the assembly [22-24, 57].

Development of the 67-nm banding pattern: The appearance of a periodic banded structure happened in the next stage, in a gradual process that started with the formation of bands 67 nm apart delineating the D-periodicity but with no further division of this period into a gap and overlap region. This division occurred later and gradually, as demonstrated by the appearance of the phosphotungstate staining bands. As the self-assembly progressed, they increased in intensity, becoming narrower and better defined, reflecting the clustering of positively and negatively charged amino acids in well-defined zones [48]. The FWHM values of the individual peaks in the mass density profiles also show that the D-band became better resolved with time:after 15, 20, and 40 minutes, the FWHM of the 12th peak (band c2), used to measure the length of the D-band, are 5.4, 5.1, and 4.1 nm, respectively (Supporting Information S6, Fig. S6 and Table S1), which reflects a 30% increase in the resolution of the peaks. In combination with an increase in the average peak heights from 1.4 a.u., 6.1 a.u., and 7.0 a.u. (Supporting Information 6, Fig. S6 and

Table 1) with increasing time, the apparent D-band periodicity should be subject to statistical variations that decrease with time, and so the value after 40 minutes is statistically more significant, as well as naturally being more representative of mature fibrils.

The variations in the length of the D-banding, together with the increase in the packing density of the fibrils as they assembled, are similar to those reported in the collagen matrix from dentin prior to mineralization [35]. Beniash et al. observed that newly deposited collagen in the proximal zone of predentin were loosely packed and displayed a periodicity of 23.5 nm with no distinguishable fine structure as found in mature collagen. Fibrils at the central zone of predentin were thicker and more closely packed, with an average D-banding of 67 nm. These similarities indicate that as collagen fibrils are deposited in the extracellular matrix of tissues, they undergo a comparable maturation process involving the gradual organization of the molecules to give rise to the D-periodicity. One important difference, however, is that collagen deposition in vivo occurs at higher concentrations in compartmentalized intracellular spaces in the presence of regulatory proteins and under temporal control over the secretion of the molecules involved[58]. Therefore, other mechanisms may also be controlling fibrillogenesis in biological tissues. It is also noteworthy that mature fibrils can still undergo alterations in their structure. Intrafibrillar mineralization in collagen during bone and dentin formation, for instance, results in changes in the molecular packing and a decrease in the length of the D-banding [36, 59].

An interesting structural feature of type I collagen fibrils of most connective tissues is that their constituent filaments are twisted with respect to the fibril axis by an angle of 17 ° [60, 61], forming a helical structure. The pitch of the twist is of several hundreds of nanometers, and was not discernible in our measurements of immature and mature fibrils. Thus, we cannot speculate at

which stage, and how, this twisting occurs as the filaments self-organize into fibrils and the collagen banding pattern arises.

Combining cryoTEM with chemical staining was critical to understand how the quarter-stagger ordering of microfibrils within a fibril developed. The first bands to appear, delineating the 67-nm repeat, correspond to the N-terminal region of the molecules. We therefore hypothesize that that the first interactions between adjacent molecules, leading to their alignment, occur through the N-terminal telopeptide. Although this hypothesis needs to be tested experimentally, it is in accordance with previous reports.Collagen assembly experiments in vitro showed that the N-terminal telopeptide is important for lateral growth of the fibril [13]. Furthermore, the binding of the N-telopeptide to a specific domain in the triple helix has been shown to induce an ordered conformation in the telopeptide, and would place the length of the D-band in the range of 66-68 nm [62], as observed in our experiments. The alignment between adjacent molecules then propagates, driven by electrostatic interactions between the positively and negatively charged amino acids in adjacent filaments. The C-terminal telopeptide region of the fibril was the last one to align, as the a2, a3 and a4 bands could only be resolved after 40 minutes. This telopeptide has been reported to be essential in the early stages of assembly, i.e., during the lag phase [13]. It is conceivable, therefore, that it mediates the association of the molecules during the formation of filaments at the early stages, and is less critical for their lateral association and further alignment to form mature fibrils. Our observations that atelo-collagen did not form fibrils support the role of the telopeptides in mediating collagen self-assembly. It is noteworthy that even fibrils that already have well-developed gap and overlap regions still display a thin, disordered layer in their periphery, which is composed of newly fused microfibrils, not aligned with the bulk of the fibril. The presence of

such disordered surface layer in collagen fibrils is similar to an amorphous layer that is present in crystalline materials [63, 64].

CONCLUSIONS

In conclusion, we have demonstrated that the self-assembly of type-I collagen molecules into fibrils in solution occurs in several stages: it starts with the formation of dimers/trimers and then thin (4-nm wide) filaments, which self-organize into fibrils by lateral aggregation, leading to the formation of disordered, loosely packed arrays of uniaxially oriented strands. The packing density of these arrays then increases, generating compact fibrils that do not yet display the D-banding periodicity. The final step is the gradual alignment of the filaments within the fibrils to form the 67-nm D-bands that are characteristic of type-I collagen. The formation of the 4 nm-thick filaments in the first step happens during the lag phase, and the subsequent stages occur in the exponential phase. Our results provide significant insights in the mechanisms of collagen self-assembly, and will help us to understand the formation of collagenous tissues in the body, such as cornea, bone, skin, and tendons, as well as elucidate how mutations in the amino-acid sequence affect fibril structure.

ASSOCIATED CONTENT

Supporting Information

Supplementary cryoTEM images, mass-density profiles, analysis of cluster-growth kinetics, mass-density analysis of the profiles of stained collagen.

Corresponding Author

*Dr. Fabio Nudelman, e-mail: fabio.nudelman@ed.ac.uk

ORCID

Fabio Nudelman: 0000-0001-7464-4309

Present Addresses

† JEOL (UK) Ltd., Silver Court, Welwyn Garden City, AL7 1LT, United Kingdom

§ Diamond Light Source, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK

Author Contributions

A.R.M., K.S.W.H and B.M. performed the cryoTEM measurements. J.O.S. and P.J.C. performed the molecular dynamics simulations. N.A.J.M. contributed to the cryoTEM measurements. F.N. designed and supervised the study. N.A.J.M, P.J.C. and F.N. co-wrote the manuscript. All authors have given approval to the final version of the manuscript. ‡ ARM and KSWH contributed equally.

ACKNOWLEDGEMENT

This work was supported by the University of Edinburgh and BBSRC, grant no. BB/M029611/1. The cryoEM facility at the University of Edinburgh was set up set up with funding from Wellcome Trust (087658/Z/08/Z) and SULSA. We thank the Center of Multiscale Electron Microscopy of Eindhoven University of Technology for access to their cryoTEM facilities.

REFERENCES

1. Orgel, J. P. R. O.; Irving, T. C.; Miller, A.; Wess, T. J., Microfibrillar structure of type I collagen in situ. *Proc Natl Acad Sci USA* **2006,** 103, (24), 9001-9005.

2. Hodge, A. J.; Petruska, J. A., Recent studies with the electron microscope on ordered aggregates of the tropocollagen macromolecule. In *Aspects of Protein Structure*, Ramachandran, G. N., Ed. Academic Press: New York, 1963; pp 289-300.

3. Orgel, J. P. R. O.; Miller, A.; Irving, T. C.; Fischetti, R. F.; Hammersley, A. P.; Wess, T. J., The in situ supermolecular structure of type I collagen. *Structure* **2001,** 9, (11), 1061-1069.

4. Comper, W. D.; Veis, A., Mechanism of Nucleation for Invitro Collagen Fibril Formation. *Biopolymers* **1977,** 16, (10), 2113-2131.

5. Comper, W. D.; Veis, A., Characterization of Nuclei in Invitro Collagen Fibril Formation. *Biopolymers* **1977,** 16, (10), 2133-2142.

6. Gelman, R. A.; Piez, K. A., Collagen Fibril Formation Invitro - a Quasi-Elastic Light-Scattering Study of Early Stages. *J Biol Chem* **1980,** 255, (17), 8098-8102.

7. Gelman, R. A.; Poppke, D. C.; Piez, K. A., Collagen Fibril Formation Invitro - Role of the Non-Helical Terminal Regions. *J Biol Chem* **1979,** 254, (22), 1741-1745.

8. Giraud-Guille, M. M., Liquid crystallinity in condensed type I collagen solutions. A clue to the packing of collagen in extracellular matrices. *J Mol Biol* **1992,** 224, (3), 861-73.

9. Gobeaux, F.; Mosser, G.; Anglo, A.; Panine, P.; Davidson, P.; Giraud-Guille, M. M.; Belamie, E., Fibrillogenesis in dense collagen solutions: a physicochemical study. *J Mol Biol* **2008,** 376, (5), 1509-22.

10. Doyle, B. B.; Hulmes, D. J. S.; Miller, A.; Parry, D. A. D.; Piez, K. A.; Woodhead.J, Axially Projected Collagen Structures. *Proc R Soc Ser B-Bio* **1974,** 187, (1086), 37-46.

11. Hulmes, D. J. S.; Miller, A.; Parry, D. A. D.; Piez, K. A.; Woodhead.J, Analysis of Primary Structure of Collagen for Origins of Molecular Packing. *J Mol Biol* **1973,** 79, (1), 137-148.

12. Helseth, D. L.; Veis, A., Collagen Self-Assembly Invitro - Differentiating Specific Telopeptide-Dependent Interactions Using Selective Enzyme Modification and the Addition of Free Amino Telopeptide. *J Biol Chem* **1981,** 256, (14), 7118-7128.

13. Prockop, D. J.; Fertala, A., Inhibition of the self-assembly of collagen I into fibrils with synthetic peptides - Demonstration that assembly is driven by specific binding sites on the monomers. *J Biol Chem* **1998,** 273, (25), 15598-15604.

14. Helseth, D. L.; Lechner, J. H.; Veis, A., Role of the Amino-Terminal Extra-Helical Region of Type-I Collagen in Directing the 4d Overlap in Fibrillogenesis. *Biopolymers* **1979,** 18, (12), 3005-3014.

15. Capaldi, M. J.; Chapman, J. A., The C-Terminal Extra-Helical Peptide of Type-I Collagen and Its Role in Fibrillogenesis Invitro. *Biopolymers* **1982,** 21, (11), 2291-2313.

16. Harris, J. R.; Reiber, A., Influence of saline and pH on collagen type I fibrillogenesis in vitro: Fibrit polymorphism and colloidal gold labelling. *Micron* **2007,** 38, (5), 513-521.

17. Li, Y. P.; Asadi, A.; Monroe, M. R.; Douglas, E. P., pH effects on collagen fibrillogenesis in vitro: Electrostatic interactions and phosphate binding. *Mat Sci Eng C-Bio S* **2009,** 29, (5), 1643-1649.

18. Na, G. C.; Phillips, L. J.; Freire, E. I., Invitro Collagen Fibril Assembly - Thermodynamic Studies. *Biochemistry-Us* **1989,** 28, (18), 7153-7161.

19. Kadler, K. E.; Hojima, Y.; Prockop, D. J., Assembly of Type-I Collagen Fibrils Denovo - between 37-Degrees and 41-Degrees-C the Process Is Limited by Micro-Unfolding of Monomers. *J Biol Chem* **1988,** 263, (21), 10517-10523.

20. Harris, J. R.; Soliakov, A.; Lewis, R. J., In vitro fibrillogenesis of collagen type I in varying ionic and pH conditions. *Micron* **2013,** 49, 60-68.

21. Jiang, F. Z.; Horber, H.; Howard, J.; Muller, D. J., Assembly of collagen into microribbons: effects of pH and electrolytes. *J Struct Biol* **2004,** 148, (3), 268-278.

22. Leow, W. W.; Hwang, W., Epitaxially Guided Assembly of Collagen Layers on Mica Surfaces. *Langmuir* **2011,** 27, (17), 10907-10913.

23. Loo, R. W.; Goh, M. C., Potassium Ion Mediated Collagen Microfibril Assembly on Mica. *Langmuir* **2008,** 24, (23), 13276-13278.

24. Narayanan, B.; Gilmer, G. H.; Tao, J. H.; De Yoreo, J. J.; Ciobanu, C. V., Self-Assembly of Collagen on Flat Surfaces: The Interplay of Collagen-Collagen and Collagen-Substrate Interactions. *Langmuir* **2014,** 30, (5), 1343-1350.

25. Gelman, R. A.; Williams, B. R.; Piez, K. A., Collagen Fibril Formation - Evidence for a Multistep Process. *J Biol Chem* **1979,** 254, (1), 180-186.

26. Bernengo, J. C.; Herbage, D.; Marion, C.; Roux, B., Inter-Molecular Interaction Studies on Native and Enzyme-Treated Acid-Soluble Collagen. *Biochim Biophys Acta* **1978,** 532, (2), 305-314.

27. Silver, F. H.; Langley, K. H.; Trelstad, R. L., Type-I Collagen Fibrillogenesis - Initiation Via Reversible Linear and Lateral Growth Steps. *Biopolymers* **1979,** 18, (10), 2523-2535.

28. Ward, N. P.; Hulmes, D. J. S.; Chapman, J. A., Collagen Self-Assembly Invitro - Electron-Microscopy of Initial Aggregates Formed during the Lag Phase. *J Mol Biol* **1986,** 190, (1), 107-112.

29. Christiansen, D. L.; Huang, E. K.; Silver, F. H., Assembly of type I collagen: fusion of fibril subunits and the influence of fibril diameter on mechanical properties. *Matrix Biol* **2000,** 19, (5), 409-420.

30. Cisneros, D. A.; Hung, C.; Franz, C. A.; Muller, D. J., Observing growth steps of collagen self-assembly by time-lapse high-resolution atomic force microscopy. *J Struct Biol* **2006,** 154, (3), 232-245.

31. LAMMPS Molecular Dynamics Simulator, URL: https://lammps.sandia.gov, accessed 22 June 2020.

32. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J Comput Phys* **1995,** 117, (1), 1-19.

33. Friedrich, H.; Frederik, P. M.; de With, G.; Sommerdijk, N. A., Imaging of self-assembled structures: interpretation of TEM and cryo-TEM images. *Angew Chem Int Ed Engl* **2010,** 49, (43), 7850-8.

34. Silver, F. H., Type-I Collagen Fibrillogenesis Invitro - Additional Evidence for the Assembly Mechanism. *J Biol Chem* **1981,** 256, (10), 4973-4977.

35. Beniash, E.; Traub, W.; Veis, A.; Weiner, S., A transmission electron microscope study using vitrified ice sections of predentin: Structural changes in the dentin collagenous matrix prior to mineralization. *J Struct Biol* **2000,** 132, (3), 212-225.

36. Quan, B. D.; Sone, E. D., Structural changes in collagen fibrils across a mineralized interface revealed by cryo-TEM. *Bone* **2015,** 77, 42-49.

37. Hyeon, C.; Thirumalai, D., Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* **2011,** 2, 487.

38. Forrey, C.; Muthukumar, M., Langevin dynamics simulations of genome packing in bacteriophage. *Biophys J* **2006,** 91, (1), 25-41.

39. Abeln, S.; Vendruscolo, M.; Dobson, C. M.; Frenkel, D., A simple lattice model that captures protein folding, aggregation and amyloid formation. *PLoS One* **2014,** 9, (1), e85185.

40. Wu, C.; Shea, J. E., Coarse-grained models for protein aggregation. *Curr Opin Struct Biol* **2011,** 21, (2), 209-20.

41. Knotts, T. A. t.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J., A coarse grain model for DNA. *J Chem Phys* **2007,** 126, (8), 084901.

42. Marenduzzo, D.; Bhattacharjee, S. M.; Maritan, A.; Orlandini, E.; Seno, F., Dynamical scaling of the DNA unzipping transition. *Phys Rev Lett* **2002,** 88, (2), 028102.

43. Pellarin, R.; Caflisch, A., Interpreting the aggregation kinetics of amyloid peptides. *J Mol Biol* **2006,** 360, (4), 882-92.

44. Nguyen, H. D.; Reddy, V. S.; Brooks, C. L., 3rd, Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett* **2007,** 7, (2), 338-44.

45. Lenoci, L.; Camp, P. J., Self-assembly of peptide scaffolds in biosilica formation: computer simulations of a coarse-grained model. *J Am Chem Soc* **2006,** 128, (31), 10111-7.

46. Chandler, D., *Introduction to Modern Statistical Mechanics*. Oxford University Press: New York, 1987.

47. Sonntag, H.; Strenge, K., *Coagulation Kinetics and Structure Formation*. Springer US: New York, 1987.

48. Chapman, J. A.; Tzaphlidou, M.; Meek, K. M.; Kadler, K. E., The Collagen Fibril - a Model System for Studying the Staining and Fixation of a Protein. *Electron Microsc Rev* **1990,** 3, (1), 143-182.

49. Meek, K. M.; Chapman, J. A.; Hardcastle, R. A., Staining Pattern of Collagen Fibrils - Improved Correlation with Sequence Data. *J Biol Chem* **1979,** 254, (21), 710-714.

50. Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G., Interplay between structure and size in a critical crystal nucleus. *Phys Rev Lett* **2005,** 94, (23), 235703.

51. De Yoreo, J. J.; Gilbert, P. U. P. A.; Sommerdijk, N. A. J. M.; Penn, R. L.; Whitelam, S.; Joester, D.; Zhang, H. Z.; Rimer, J. D.; Navrotsky, A.; Banfield, J. F.; Wallace, A. F.; Michel, F. M.; Meldrum, F. C.; Colfen, H.; Dove, P. M., Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* **2015,** 349, (6247), aaa6760.

52. Erdemir, D.; Lee, A. Y.; Myerson, A. S., Nucleation of crystals from solution: classical and two-step models. *Acc Chem Res* **2009,** 42, (5), 621-9.

53. Fleischmajer, R.; Perlish, J. S.; Faraggiana, T., Rotary shadowing of collagen monomers, oligomers and fibrils during tendon fibrillogenesis. *J Histochem Cytochem* **1991,** 39, (1), 51-58.

54. Trelstad, R. L.; Hayashi, K.; Gross, J., Collagen fibrillogenesis: Intermediate aggregates and suprafibrillar order. *Proc Natl Acad Sci USA* **1976,** 73, (11), 4027-4031.

55. Camp, P. J.; Mason, C. P.; Allen, M. P.; Khare, A. A.; Kofke, D. A., The isotropic-nematic phase transition in uniaxial hard ellipsoid fluids: Coexistence data and the approach to the Onsager limit. *J Chem Phys* **1996,** 105, (7), 2837-2849.

56. Onsager, L., The Effects of Shape on the Interaction of Colloidal Particles. *Ann Ny Acad Sci* **1949,** 51, (4), 627-659.

57. Fang, M.; Goldstein, E. L.; Matich, E. K.; Orr, B. G.; Holl, M. M. B., Type I Collagen Self-Assembly: The Roles of Substrate and Concentration. *Langmuir* **2013,** 29, (7), 2330-2338.

58. Kadler, K. E.; Hill, A.; Canty-Laird, E. G., Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. *Current Opinion in Cell Biology* **2008,** 20, 495-501.

59. Fratzl, P.; Fratzl-Zelman, N.; Klaushofer, K., Collagen Packing and Mineralization - an X-Ray-Scattering Investigation of Turkey Leg Tendon. *Biophys J* **1993,** 64, (1), 260-266.

60. Ottani, V.; Martini, D.; Franchi, M.; Ruggeri, A.; Raspanti, M., Hierarchical structures in fibrillar collagens. *Micron* **2002,** 33, (7-8), 587-96.

61. Ottani, V.; Raspanti, M.; Ruggeri, A., Collagen structure and functional implications. *Micron* **2001,** 32, (3), 251-60.

62. Malone, J. P.; George, A.; Veis, A., Type I collagen N-telopeptides adopt an ordered structure when docked to their helix receptor during fibrillogenesis. *Proteins* **2004,** 54, (2), 206-215.

63. Huang, S. J.; Tsai, Y. L.; Lee, Y. L.; Lin, C. P.; Chan, J. C. C., Structural Model of Rat Dentin Revisited. *Chem Mater* **2009,** 21, (13), 2583-2585.

64. Nassif, N.; Pinna, N.; Gehrke, N.; Antonietti, M.; Jager, C.; Colfen, H., Amorphous layer around aragonite platelets in nacre. *Proc Natl Acad Sci USA* **2005,** 102, (36), 12653-12655.

For Table of Content use only

Disordered Filaments Mediate the Fibrillogenesis of Type-I Collagen in Solution

Andrew McCluskey, Kennes S. W. Hung, Bartosz Marzec, Julien O. Sindt, Nico A. J. M. Sommerdijk, Philip J. Camp, and Fabio Nudelman