



Sequence-specific response of collagen-mimetic peptides to osmotic pressure

Lorena Ruiz-Rodriguez,¹ Philip Loche,¹ Lise Thornfeldt Hansen, Roland R. Netz,¹ Peter Fratzl,¹ Emanuel Schneck,^{2*} Kerstin G. Blank,^{2*} and Luca Bertinetti^{3*}

Collagen is an essential protein in mammalian extracellular tissues and a better understanding of its mechanical function is important both from a materials science and from a biomedical viewpoint. Recently, collagen has been shown to contract along the fibre direction when subjected to osmotic stress, a process that could play important roles in strengthening bone and in developing tissue tension during extracellular matrix development. The present work uses collagen-like short peptides to show that the canonical collagen sequence is not responsible for this contraction. The conclusion is that the collagen amino acid sequence must have evolved to include guest sequences within the canonical glycine-proline-hydroxyproline repeat that provide the observed contractility.

Native collagen molecules usually contract upon dehydration, but the details of their interaction with water are poorly understood. Previous molecular modeling studies indicated a spatially inhomogeneous response, with a combination of local axial expansion and contraction. Such sequence-dependent effects are difficult to study with native collagen. In this article, we use collagen-mimetic peptides (CMPs) to investigate the effect of osmotic pressure on several collagen-mimetic sequences. Synchrotron x-ray diffraction combined with molecular dynamics simulations shows that CMPs pack differently depending on osmotic pressure and exhibit changes in the helical rise per residue of individual molecules. Infrared spectroscopy reveals that osmotic pressure affects the stability of the triple helix through changes in triple helix-stabilizing hydrogen bonds. Surprisingly, CMPs with the canonical collagen sequence glycine-proline-hydroxyproline are found to elongate upon dehydration, while sequence modifications are able to reverse this tendency. This strongly suggests that the overall contraction of native collagen molecules is not programmed into the canonical sequence but is specific to local amino acids that substitute for proline or hydroxyproline along the protein chain.

Introduction

Collagen, the most abundant protein in mammals, plays a key structural and biochemical role in the extracellular matrix.^{1,2} All 28 different collagen types known to date contain at least one right-handed helical domain, made up of three polypeptide chains with the repetitive sequence

(Xaa-Yaa-Gly)_n. In mammalian collagens, Pro is frequently found in the Xaa position, while the Yaa position is often occupied by the post-translationally modified Pro derivative 4*R*-hydroxy-2*S*-proline (Hyp). The triple helical structure, which is stabilized by one interchain hydrogen

Lorena Ruiz-Rodriguez, Mechano(bio)chemistry, Max Planck Institute of Colloids and Interfaces, Germany; Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Germany; jlarena2r@gmail.com
 Philip Loche, Department of Physics, Freie Universität Berlin, Germany; ploche@physik.fu-berlin.de
 Lise Thornfeldt Hansen, Department of Physics, Freie Universität Berlin, Germany; lthornfeldt@gmail.com
 Roland R. Netz, Department of Physics, Freie Universität Berlin, Germany; rnetz@physik.fu-berlin.de
 Peter Fratzl, Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Germany; Peter.Fratzl@mpikg.mpg.de
 Emanuel Schneck, Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Germany; Department of Physics, Technische Universität Darmstadt, Germany; schneck@fkp.tu-darmstadt.de
 Kerstin G. Blank, Mechano(bio)chemistry, Max Planck Institute of Colloids and Interfaces, Germany; kerstin.blank@mpikg.mpg.de
 Luca Bertinetti, Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Germany; B CUBE - Center for Molecular Bioengineering, Technische Universität Dresden, Germany; luca.bertinetti@tu-dresden.de
 *Corresponding authors
 doi:10.1557/s43577-021-00138-9



bond per Xaa-Yaa-Gly tripeptide unit,^{3,4} is highly abundant in fibril-forming collagens. These include collagen I, which is an important building block of connective tissues, such as tendons, ligaments, dermis, cornea, and bone.^{1,2} Upon fibrillation, triple helices self-assemble into a highly regular, periodic structure where molecules are axially staggered with a so-called D-period of 67 nm.^{5–7} This staggered alignment results in tightly packed overlap regions and less dense gap regions. In the overlap regions, the triple helices are laterally packed in a distorted hexagonal fashion, while the molecular arrangement in the gap regions is more variable.^{6–9}

Many studies of fibrillar collagens and triple helix-forming collagen-mimetic peptides (CMPs) have shown that water content (i.e., osmotic pressure) plays a crucial role in determining collagen structure and function.^{4,10–19} Hydration and protein-induced water structuring are fundamental processes that determine the biochemical properties of the triple helical collagen molecule and its higher-order assemblies as well as its tightly connected multiscale mechanical performance. For example, small-angle x-ray scattering (SAXS) and Raman experiments on collagen-based tissues have shown that small changes in osmotic pressure significantly affect the lateral packing density of collagen triple helices and their interaction with water.^{20,21} Atomic force microscopy of collagen fibrils has further shown higher degrees of osmotic pressure-induced swelling in the gap regions when compared to the overlap regions.²² Osmotic pressure-dependent, dynamic structural features thus critically determine solute diffusion, exposure of integrin binding and protease cleavage sites,^{19,23} accessibility of nucleation sites for mineralization⁹ as well as the tensile and compressive stiffness along and perpendicular to the fibril axis.^{22,24–26}

For fibers of collagen I, it has been shown that osmotic pressure-dependent mechanical properties are directly related to molecular conformational changes that affect the axial length.²⁷ Within the physiological osmotic pressure range, this length change can generate axial forces larger than those produced by muscles. In mineralized collagen-based tissues, these forces further induce enormous compressive stresses (estimated to reach 1 GPa) that act on the apatite crystals associated with the collagen fibrils.²⁵ This purely passive physicochemical mechanism is fundamentally different from other biological processes where force generation requires energy in the form of ATP (e.g., as in the case of molecular motors). Molecular dynamics (MD) simulations suggested that these conformational changes are not homogeneously distributed along the collagen molecule.²⁷ Instead, some regions elongated, while others contracted so that the net molecular contraction observed upon dehydration (and vice versa in the case of rehydration) was predicted to be the sum of highly localized processes. The axial response of collagen to osmotic pressure is thus expected to be finely controlled by the local amino acid sequence. Sequence-level control over the macroscopic hydration-induced strain also exists in silk;^{28,29} however, silk is based on β -sheets rather than helical conformations

and contracts upon hydration. Other, fundamentally different osmotic pressure responses in biochemical systems include water-based actuation in plants³⁰ and bacillus spores.³¹

To understand the origin and functional consequences of local hydration patterns in collagen I, it is essential to investigate the structure of individual triple helices and their higher-order assemblies in a sequence-controlled manner. Each triple helix of full-length, fibrillar collagen I consists of a large number of amino acids (~ 3000), forming more than 200 unique Xaa-Yaa-Gly tripeptide units.³² It is thus not easily possible to obtain sequence-controlled and thermodynamically stable fragments of native collagen and to investigate their properties as a function of environmental changes. To overcome this limitation and to obtain precise structural and chemical information at the molecular level, we utilized CMPs and investigated their response to osmotic pressure changes. CMPs are chemically well-defined, synthetic peptides with the sequence (Xaa-Yaa-Gly)_n that fold into the characteristic triple helical structure.^{33–36} First CMPs primarily consisted of the canonical sequences (Pro-Pro-Gly)_n and (Pro-Hyp-Gly)_n.^{37,38} These CMPs have critically contributed to understanding the fine molecular differences determining collagen structure and hydration patterns,^{16,39–44} folding,^{45–47} thermodynamic stability,^{37,38,46,48–50} and higher-order assembly⁵¹ as well as the specific role of Hyp in determining the above-mentioned properties.^{16,38,40,49–52} Later, so-called host-guest CMPs have been introduced, where either individual amino acids or tripeptide units were modified in the central part of the peptide sequence.^{13,53–55} With this strategy, also disease-related mutations have been investigated¹³ as well as segments of biochemically active sequences, such as protease cleavage and integrin binding sites.^{17,36,56,57}

In this work, we utilized CMPs to investigate the response of different sequences to changes in osmotic pressure. While precisely controlling the relative humidity (RH) of microcrystalline CMP powder samples, we determined three essential structural parameters that are critical for understanding triple helix conformational changes as well as higher-order assembly. Specifically, we performed *in situ* infrared (IR) spectroscopy to obtain information about the effect of CMP-water interactions on the strength of the interchain hydrogen bond. Synchrotron-based x-ray diffraction (XRD) was used in combination with MD simulations to investigate the lateral packing arrangement, specifically focusing on the center-to-center distance between CMP triple helices. The XRD data were further used to obtain the helical rise per residue to obtain information about the axial molecular contraction/elongation of individual triple helices. We show that decreasing the osmotic pressure causes stronger interchain hydrogen bonding. Furthermore, the chemical properties of the triple helix surface influence the swelling behavior (lateral packing) of higher-order triple helix assemblies, causing a sequence-specific rearrangement of the helices in the range of physiologically relevant osmotic pressures. The peptide sequence also critically determines the osmotic pressure-dependent helical rise per residue and even



small changes in the sequence can reverse the direction of hydration-induced axial deformation. These results provide support for the hypothesis that the response of full-length collagen to osmotic pressure is indeed controlled at the level of the local amino acid sequence.

Results

Sequence design

With the goal of investigating the sequence-specific response of CMPs to changes in osmotic pressure, we utilized four different CMPs, each consisting of 10 tripeptide units. This CMP length ensures sufficient thermodynamic stability,^{38,46,49,50,58,59} is known to yield well aligned triple helical structures and has frequently been used for crystal structure analysis.^{11,13,16,39–43,59} Using the canonical sequences (Pro-Pro-Gly)₁₀ and (Pro-Hyp-Gly)₁₀, abbreviated as (PPG)₁₀ and (POG)₁₀ in the following, we first focused on the ability of the imino acids to participate in hydrogen bonds. The tripeptide units PPG and POG occur in human collagen I with a frequency of 1% and 9.6%, respectively (**Figure S1**). These two CMPs were complemented with the peptide (Hyp-Hyp-Gly)₁₀, abbreviated as (OOG)₁₀. The OOG tripeptide unit is not a natural building block of

mammalian collagens and has frequently been used for investigating the role of Hyp in determining collagen structure and stability.^{59–63}

The structures of these three CMPs at low osmotic pressure have previously been determined with x-ray crystallography.^{40,42–44,59,61} They all form triple helices described by the so-called 7/2 model, with 7 amino acids per 2 turns of the helix (**Figure 1**).⁶⁴ Their crystal structures revealed that water decorates the triple helix in a highly ordered manner, forming well-defined first and second hydration shells.^{40,43,44,59} A large number of water molecules interact directly with the main chain carbonyl (C=O) groups of Gly and Yaa and occupy identical positions in (PPG)₁₀, (POG)₁₀ and (OOG)₁₀; however, also clear variations in the hydration patterns have been established for different CMPs. Specifically, hydroxyl groups of Hyp serve as anchoring points for the water network surrounding the triple helix while clathrate-like water structures assemble around hydrophobic Pro residues.^{14,62} Several crystal structures have further revealed direct hydrogen bonds between neighboring triple helices, with their number and location being dependent on the specific structure of the CMP and the surrounding water network.^{43,44,59} Interestingly, localized changes in hydration were found for host–guest CMPs,

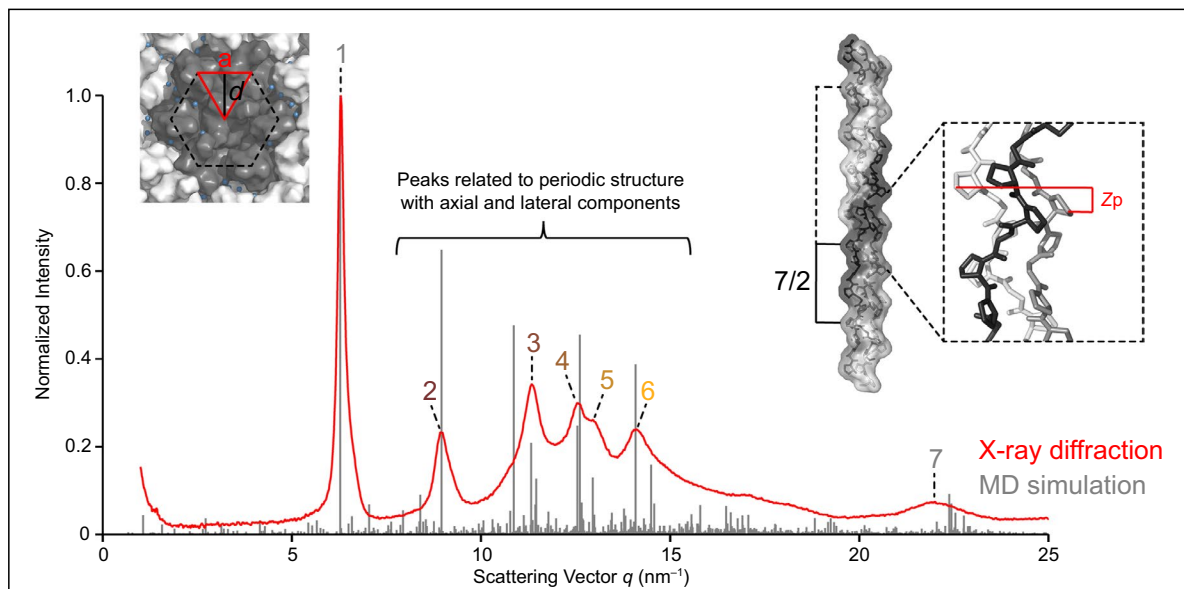


Figure 1. Scattering profile $S(q)$ of (PPG)₁₀ in dry conditions. Shown is an experimentally obtained $S(q)$ profile with the assignment of peaks 1–7 (RH \approx 10%; $\Pi \approx$ 320 MPa). Peak 1 with a Bragg distance of $d_1 = 1.0$ nm is related to the lateral center-to-center distance a (top left inset). Peak 7 represents the helical rise per residue z_p (molecular axial model on the right). Several of the peaks 2–6 originate from scattering units located in lattice planes that combine lateral and axial components. For the molecular dynamics (MD) simulation, an array of 8×8 (PPG)₇ triple helices was used. Considering the 7/2 model for CMPs, (PPG)₇ was repeated periodically in the axial direction to obtain infinitely long triple helices. From the simulation containing 0.4 water molecules per amino acid, discrete scattering line intensities were computed for each accessible set of Miller indices (h, k, l) , as described in the section on “Materials and Methods.” To account for differences in the hydration level, the q -axis associated with the simulated data was rescaled by a factor 1.041 such that the position of peak 1 matches the experimental value. For better comparison, the intensities were normalized with respect to peak 1. Calculated lines with the highest relative intensities can be assigned to the experimentally detected peaks (**Table S1**).

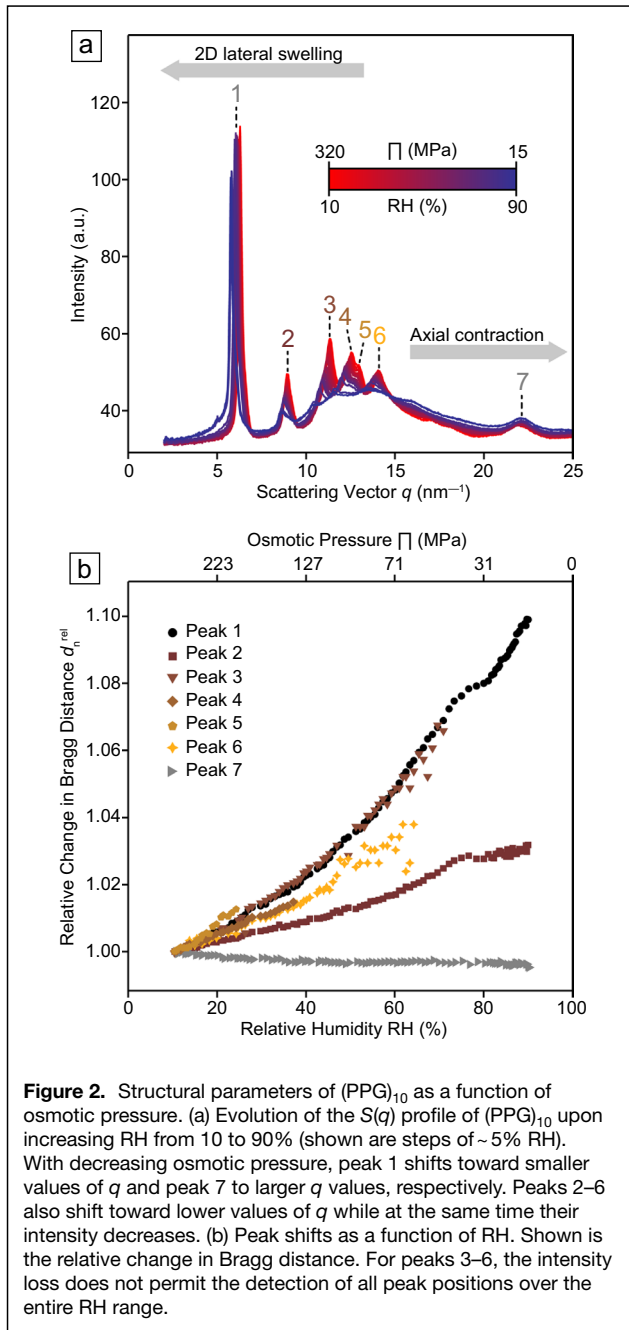


Figure 2. Structural parameters of (PPG)₁₀ as a function of osmotic pressure. (a) Evolution of the $S(q)$ profile of (PPG)₁₀ upon increasing RH from 10 to 90% (shown are steps of ~5% RH). With decreasing osmotic pressure, peak 1 shifts toward smaller values of q and peak 7 to larger q values, respectively. Peaks 2–6 also shift toward lower values of q while at the same time their intensity decreases. (b) Peak shifts as a function of RH. Shown is the relative change in Bragg distance. For peaks 3–6, the intensity loss does not permit the detection of all peak positions over the entire RH range.

where specific Pro and Hyp residues were substituted with other amino acids.^{17,19} The fourth CMP studied here contains the guest sequence Ala-Arg-Gly-Ser-Asp-Gly inserted between two (Pro-Pro-Gly)₄ host sequences, so that the full sequence is (Pro-Pro-Gly)₄-Ala-Arg-Gly-Ser-Asp-Gly-(Pro-Pro-Gly)₄. This specific guest sequence is found in the $\alpha 2$ chain of collagen I (Figure S1) and is highly conserved among mammalian species. Using MD simulations, this sequence was predicted to be highly sensitive to osmotic pressure, displaying significant contraction upon dehydration.²⁷

Peak assignment of XRD scattering patterns

To quantify osmotic pressure-induced structural changes, we performed *in situ* XRD experiments. We typically started with a powder of microcrystalline CMPs, obtained by freeze drying, that was equilibrated at low relative humidities (RH < 5%). Subsequently, RH was slowly increased in a controlled fashion and the scattered x-ray intensity was collected on a 2D detector (Figure S2A).

Typically, the scattering profiles consisted of concentric rings of azimuthally isotropic intensities. This confirms that the CMP powder samples consist of small, randomly oriented domains of coherently aligned triple helical structures. For (PPG)₁₀ measured in dry conditions (RH ≈ 10%; $\Pi \approx 320$ MPa), the azimuthally integrated 2D scattering profile (I versus q plot; $S(q)$) exhibits several peaks (named peak 1 to 7, located at q_1 – q_7) (Figure 1). Each peak contains information about the characteristic real-space distance correlations within the system that can be calculated as $d_n = 2\pi/q_n$. The measured distances indicate correlations over a larger range than only the very first neighbors. The integrated profile is not straightforward to interpret as the measured $S(q)$ profile is not compatible with the diffraction pattern of highly hydrated crystal structures and no reference crystal structure is available for high osmotic pressures.

Based on earlier experimental work on dehydrated collagen fibers,^{21,65} we initially consider a hexagonal arrangement of triple helices (Figure 1). In this case, the $S(q)$ profile carries information about the Bragg distance between triple helices d_1 (peak located at $q_1 \approx 6$ nm⁻¹, corresponding to a distance $d_1 \approx 1.0$ nm), which in turn yields the center-to-center distance $a = d_1/(\sqrt{3}/2)$. The $S(q)$ profile further shows the helical rise per residue z_p that describes the spacing between amino acids in the axial direction (peak located at $q_7 \approx 22$ nm⁻¹, corresponding to a distance $d_7 \approx 0.285$ nm; Figure 1). The positions of peaks 1 and 7 closely resemble those of natural collagen.^{25,66}

Most of the other characteristic peaks, located between $q = 7$ nm⁻¹ and $q = 15$ nm⁻¹, cannot be directly related to a 2D hexagonal arrangement of objects. This becomes evident when following the relative peak shifts as a function of osmotic pressure (Figure 2). Upon hydration, all peaks shift to smaller values of q (except peak 7, which is related to z_p); however, the extent of this change in the relative Bragg distance ($d_n^{rel} = d_n(RH)/d_n^{dry}$) is different (Figure 2b). Peaks related to the in-plane arrangement of the triple helices are expected to fall on the same normalized curve (i.e., the curve of d_1^{rel}). As this is only the case for some peaks, we propose that the other diffraction peaks originate from lattice planes having components in both axial and lateral directions.

To validate this interpretation, complementary MD simulations were performed. The initial unit cell of the simulation contained 8×8 CMPs, each consisting of triple helix chains with 7 PPG tripeptide units (Figure S3). The 64 triple helices were then placed into a triclinic simulation box with equal spacing in the x, y directions. The triple helices were periodically repeated in the z direction and the system was hydrated



using 0.4 water molecules per amino acid (N_W/N_{AA}), corresponding to a low RH in the experiment (see also supporting information for an estimate of N_W/N_{AA} in the XRD experiment). In the experiment, the coherent scattering domains are randomly oriented (powder averaging). In contrast, all triple helices are aligned in the simulations so that all three spatial directions can be distinguished. When comparing the experimental $S(q)$ profiles with those obtained numerically from the MD simulations, it is thus possible to identify the relevant crystallographic planes associated with each experimental peak in terms of the three Miller indices (Figure 1). For this assignment, the magnitude q of the scattering vector \mathbf{q} associated with the strongest lines in the simulation-based structure factor was compared to the experimental peak positions (Table S1). The result clearly confirms that the $S(q)$ profile of (PPG)₁₀ indeed corresponds to a distorted hexagonal packing of triple helices (Figures S3–S4).

The comparison between XRD and MD results further shows that the distance d_1 , as well as d_3 and d_4 , are associated to planes parallel to the triple helix axis (i.e., $l = 0$). For these peaks, the increase of d_n^{rel} with hydration is thus expected to be identical, as is clearly observed for peaks 1 and 3 (Figure 2b). Peak 4 does not meet this expectation. This may be explained with the overlap of peaks 4 and 5, which complicates the fitting of both peaks with increasing hydration. Distances d_2 , d_5 , and d_6 correspond to planes with Miller indices (0,8,6), (8,12,6) and (16,8,6). These planes are neither parallel nor perpendicular to the axial direction. As a result, the observed increase in d_n^{rel} is smaller than for the purely lateral peaks (Figure 2b). The experimental $S(q)$ profiles of (POG)₁₀ and (OOG)₁₀ (Figure S5) also exhibit peaks 2–6 although their intensities are lower. This is a strong indication that also these CMPs assume a similar distorted hexagonal arrangement at high osmotic pressures. The lower intensities of the peaks with high Miller indices indicate that (POG)₁₀ and (OOG)₁₀ are less ordered than (PPG)₁₀.

Crystalline and partially crystalline molecular packing as a function of hydration

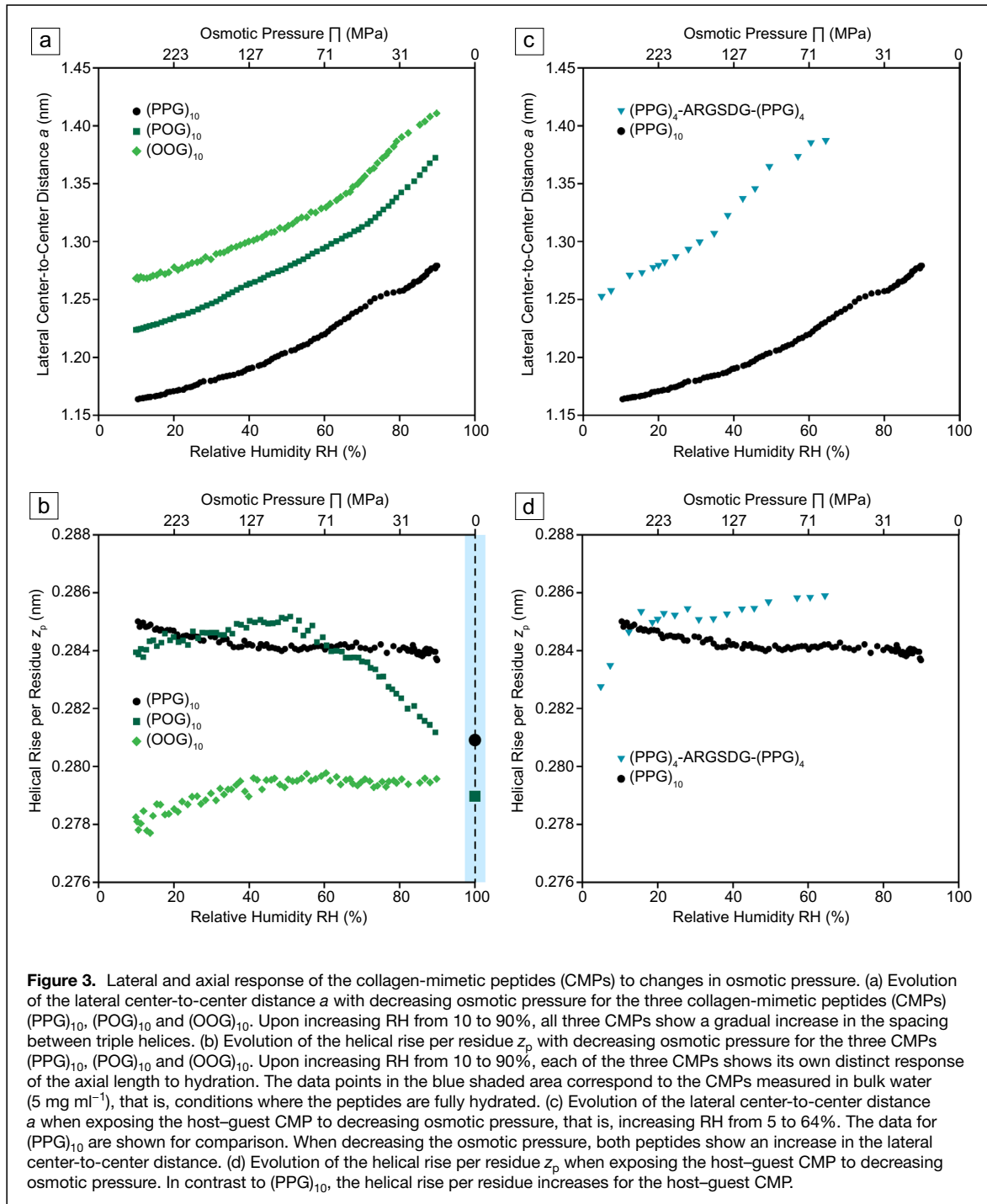
The osmotic pressure-dependent evolution of the lateral center-to-center distance was subsequently compared for the three different CMPs (PPG)₁₀, (POG)₁₀ and (OOG)₁₀ (Figure 3a). The peptides exhibit different center-to-center distances in dry conditions. The values of a are 1.16 nm for (PPG)₁₀, 1.21 nm for (POG)₁₀ and 1.26 nm for (OOG)₁₀, respectively. As hydroxylation increases the side chain volume, it directly affects the effective triple helix diameter and, as a result, the center-to-center distance between triple helices. The effective packing depends on the optimal hydrogen bond arrangements between the helices, in which the side chains may play an important role. Interestingly, (PPG)₁₀ exhibits sharper Bragg peaks than (POG)₁₀ and (OOG)₁₀ (Figure S5). This indicates higher structural order in (PPG)₁₀ crystals as compared to the other two CMPs.

When increasing RH, the center-to-center distance increases almost monotonically for all CMPs (Figure 3a). This is consistent with the incorporation of water between the triple helices; however, the increment Δa ($\Delta a = a_{>95\%} - a_{10\%}$) merely reaches ~ 0.1 – 0.2 nm at the highest value of RH that could be reached experimentally. This is less than the diameter of a water molecule (0.3 nm). This behavior clearly illustrates that the triple helices cannot be simplified as hard cylinders. If the CMPs were hard cylinders, discrete swelling steps of the size of water molecules would be expected. Instead, a smooth and gradual increase in the center-to-center distance is observed upon water uptake. This suggests that the triple helices are soft, deformable and chemically heterogeneous structures that feature niches where water molecules can insert locally. The localized binding of water causes only small changes in the molecular arrangement and gradually expands the average lateral distance between CMPs. This interpretation is clearly corroborated when comparing the experimental data to MD simulations performed at two different hydration levels (0.4 and 0.8 N_W/N_{AA}).

Figure 4a shows the radial distribution function of water molecules around (PPG)_n in cylindrical coordinates around the central axis of the triple helix (RDF_{axis}). In this representation, where the water distribution is averaged along the axial direction, it is seen that water molecules assume a rather continuous distribution between each triple helix and its direct neighbors, in line with the deformability and softness emphasized before. Upon increasing the hydration level, the positions of the maxima of the water distributions change only slightly, while the absolute RDF_{axis} values increase proportionally with the amount of water. This reflects that the water layer expands only slightly in the equatorial direction. In fact, when increasing the hydration level from 0.4 to 0.8 N_W/N_{AA} , the lateral center-to-center distance increases only by $\delta a = 0.09$ nm.

On the other hand, the spherical radial distribution function of water around the oxygen atom of the carbonyl groups ($RDF_{C=O}$), shows sharp distinct maxima (Figure 4b). This demonstrates that water interacts preferentially with determined collagen sites (i.e., the exposed C=O groups), which are the only polar groups on the outer surface of (PPG)_n. In fact, water forms a specific structure characterized by long-range order up to ~ 1 nm. When comparing $RDF_{C=O}$ at the two different hydration levels, an increase of the first peak at around 0.28 nm is observed. The increase of this peak, which corresponds to water molecules that are directly bound to the C=O groups, is smaller than the average increase seen at larger distances. This indicates that water preferentially binds within collagen niches at low hydration. With increasing hydration, more and more water molecules insert between triple helices. This localized binding can also be observed in the simulation snapshot in Figure 4b.

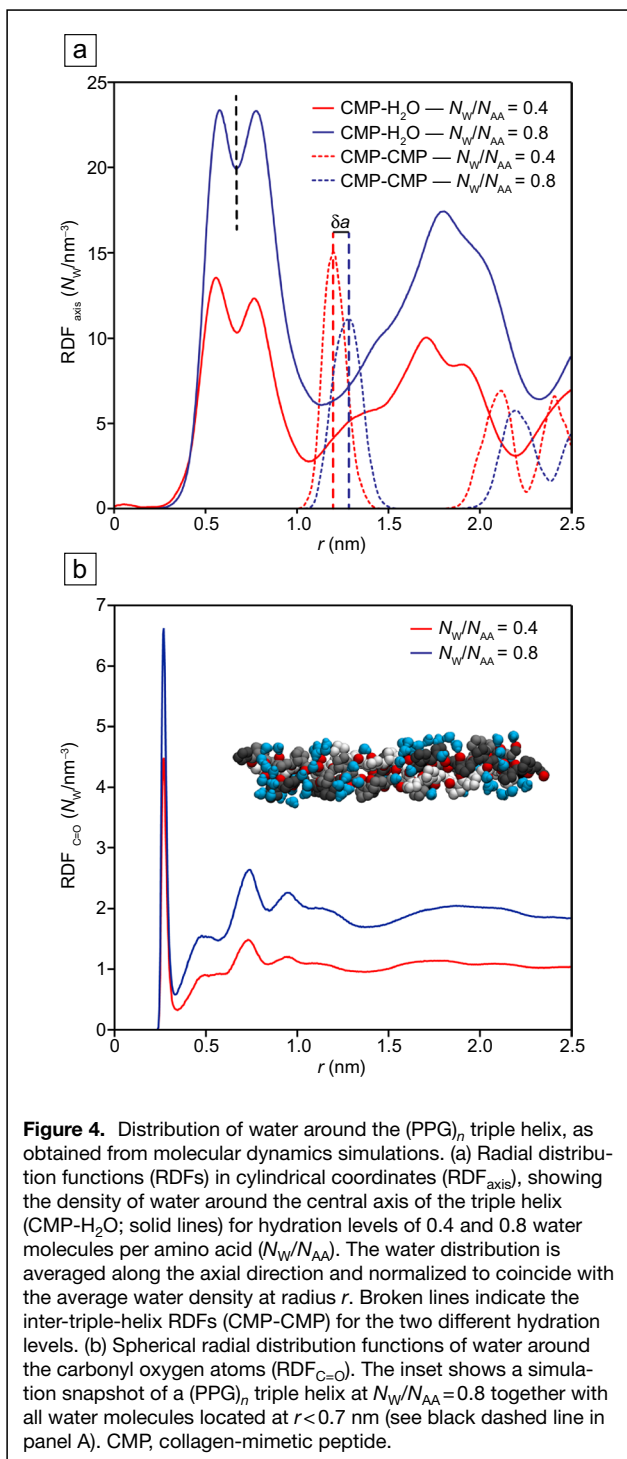
XRD experiments were also performed using CMP samples fully immersed in water, that is, at full hydration (Figure 5; Figure S6). When comparing the $S(q)$ profiles of (PPG)₁₀ obtained at full hydration with the $S(q)$ profile of the sample



with controlled RH, it is observed that the inner peak (peak 1, related to the center-to-center distance a) splits into four peaks when the CMPs have reached their highest possible level of hydration. This peak splitting originates from the rearrangement of the hexagonal packing into a new configuration with its own characteristic lateral distances. The new, hydrated configuration is nearly identical to previously reported crystal

structures for $(PPG)_{10}$.^{40,42} Excellent overlap of scattering peaks is obtained when comparing the $S(q)$ profile of our XRD results with the $S(q)$ profile obtained when rotationally averaging single crystal data of the PDB entry 1a3j;⁴⁰ (Figure 5).

The configuration found for the crystal structure corresponds to a Penrose tiling (i.e., a combination of squares and triangles). For this configuration, the lateral center-to-center



distance is ~ 1.4 nm between first neighbors and ~ 1.95 nm between second neighbors (Figure 5).⁴⁰ For our experiments, this suggests that the triple helices are arranged in a distorted hexagonal configuration at low to intermediate hydration levels (at least up to $N_W/N_{AA} = 1$), while a new configuration with fivefold in-plane coordination is obtained at higher hydration levels. The triple helices thus do not reach infinite dilution

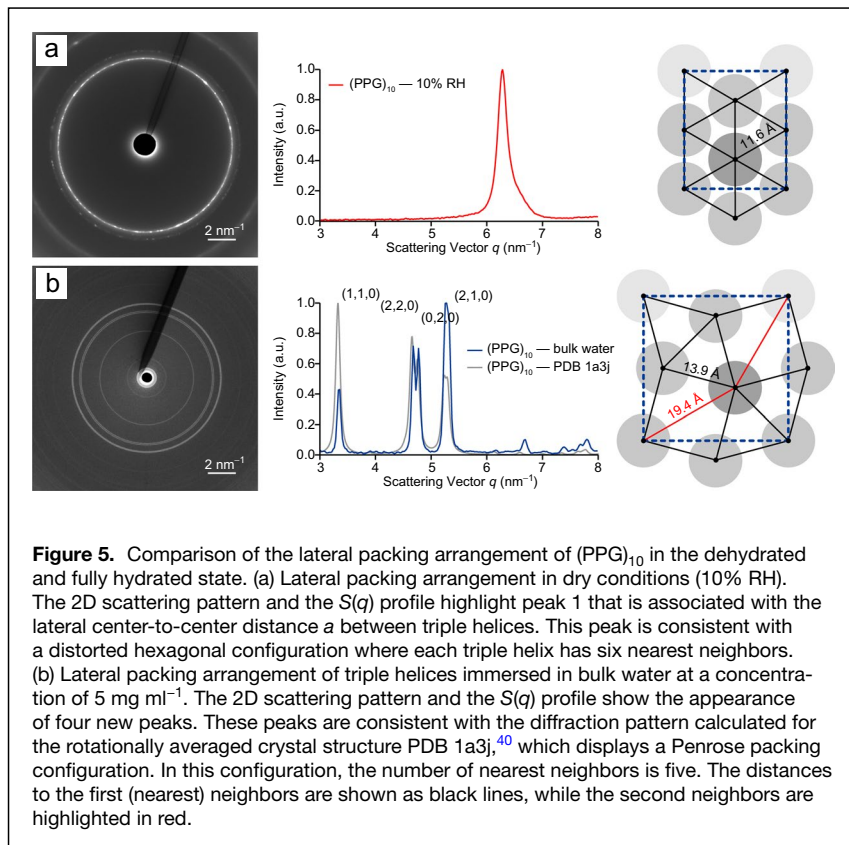
when immersed in bulk water. Instead, they organize into a crystalline assembly with localized interactions between neighboring triple helices and a specific water network with $N_W/N_{AA} \approx 2$ (Figure 5).⁴⁰

The CMP $(POG)_{10}$ shows a similar behavior as $(PPG)_{10}$. When immersed in bulk water, the scattering profile shows a reorganization into a configuration with a fourfold in-plane coordination of triple helices, as observed in the published crystal structure of $(POG)_{11}$ (PDB 1v6q;⁴³ Figure S6). This again confirms that the triple helices are not infinitely diluted in bulk water but reach a maximum hydration of $N_W/N_{AA} \approx 2$. While the coordination changed from six to five nearest neighbors for $(PPG)_{10}$, only four nearest neighbors are observed for $(POG)_{10}$. For $(OOG)_{10}$, the intensity profile in bulk water was not experimentally accessible, as the CMPs completely dissolved under these conditions. For the closely related CMPs $(OOG)_{10}$ and $(GOO)_9$, also no consistent crystal structures have been obtained, suggesting that CMPs containing OOG tripeptide units interact only weakly.^{59,61} The reduced overall strength of interhelix interactions is most likely caused by the higher number of hydrogen bond donors and acceptors even though we cannot fully exclude that it may be caused by the presence of 4*R*-hydroxy-2*S*-proline in the Xaa position of $(OOG)_{10}$ (instead of the naturally occurring 3*R*-hydroxy-2*S*-proline).

Upon increasing the hydration level to $2 N_W/N_{AA}$, the CMPs $(PPG)_{10}$ and $(POG)_{10}$ appear to display common features as well as clear sequence-specific differences. In both cases, a transition from a hexagonal packing to more complex structures is observed; however, these structures are characterized by varying numbers of nearest neighbors. This indicates that the presence and number of hydroxyl groups in the side chains dictates the nature of interhelical interactions or allows the CMP to dissolve, as we observed for $(OOG)_{10}$. For $(PPG)_{10}$ and $(POG)_{10}$, it appears more energetically favorable to break only a few interhelix bonds so that large volumes open up for the local accommodation of water molecules. This is particularly relevant for understanding previously observed differences between the gap and overlap regions and within parts of the gap regions in natural collagen I.⁹ Specifically, this observation may help to explain why the triple helices seem to show a hexagonal arrangement in the overlap regions while they lose this arrangement in some parts of the gap regions and assume a highly distorted configuration, where specific interhelical interactions are possibly maintained.

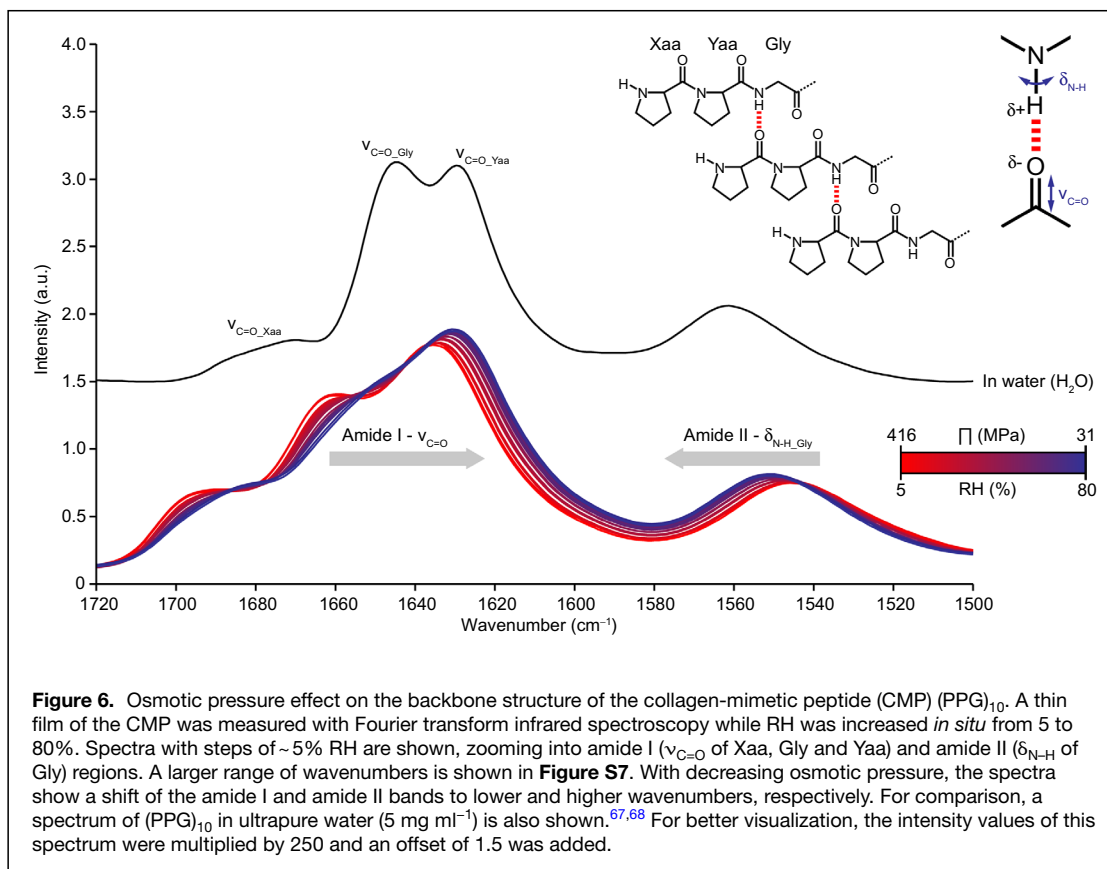
Effect of hydration on intramolecular hydrogen bonding

In addition to alterations in the lateral packing of triple helices, also osmotic pressure-dependent changes in molecular bonding and structure are expected. Vibrational spectroscopies allow for observing such changes in molecular conformation and interactions. In the case of CMPs, FTIR spectroscopy has been shown to provide information about the triple helix-stabilizing interchain hydrogen bond between the amine group



(N–H) of Gly and the carbonyl (C=O) in the Xaa position (**Figure 6**) as well as about the interaction of the C=O groups of Gly and Yaa with solvent.^{67,68} Here, FTIR spectroscopy was used to follow the molecular properties of the respective C=O and N–H bonds for (PPG)₁₀, (POG)₁₀ and (OOG)₁₀ while gradually increasing RH *in situ* (**Figure S2B**). The FTIR spectra of these CMPs exhibit a characteristic band located between 1700–1600 cm⁻¹ (**Figure 6**), which corresponds to the stretching vibration of C=O ($\nu_{C=O}$; amide I). As reported earlier,^{67,68} the components at 1670 cm⁻¹, 1645 cm⁻¹ and 1629 cm⁻¹ can be assigned to the different amino acids Xaa, Gly and Yaa, respectively. The band located between 1500 and 1580 cm⁻¹ corresponds to the bending of the Gly N–H bond ($\delta_{N-H_{Gly}}$; amide II) (**Figure 6**).

When increasing RH, all peaks related to $\nu_{C=O}$ gradually shift to lower wavenumbers. At the same time, $\delta_{N-H_{Gly}}$ shifts in the opposite direction.





The shifts of $\nu_{\text{C=O}_{\text{Gly}}}$ and $\nu_{\text{C=O}_{\text{Yaa}}}$ probably arise from stronger and stronger hydrogen bonding of the C=O groups with the surrounding solvent. The C=O groups of Gly and Yaa are located on the triple helix surface and are thus exposed to more and more water molecules when the hydration level increases (as also indicated by the MD results). Interestingly, the engagement of these C=O groups in hydrogen bonding also affects the C=O group of Xaa. Being involved in the interchain hydrogen bond, the C=O group of Xaa does not directly interact with solvent. The observed shift of $\nu_{\text{C=O}_{\text{Xaa}}}$ to lower wavenumbers thus indicates that the change in the chemical environment around the Gly and Yaa carbonyls affects the entire backbone of the molecule, causing the interchain hydrogen bond to become stronger with increasing hydration. This interpretation is supported by a shift of $\delta_{\text{N-H}_{\text{Gly}}}$ toward higher wavenumbers. As the hydrogen bond becomes stronger, the proton of the N-H bond is more strongly attracted to the hydrogen bond acceptor. This makes the N-H bond harder to bend and $\delta_{\text{N-H}}$ shifts to higher wavenumbers.

The FTIR spectra for the CMPs (POG)₁₀ and (OOG)₁₀ show similar trends even though the $\nu_{\text{C=O}}$ peaks are less well resolved (Figure S7). Altogether, these results indicate that the interaction between the triple helix and water increases the strength of the interchain hydrogen bond between the C=O group of Xaa and the N-H bond of Gly for all three CMPs. As this hydrogen bond is oriented perpendicular to the triple helix axis, we conclude that the helical structure becomes more stable at higher hydration. The degree of this stabilization appears to be different for the three CMPs (PPG)₁₀, (POG)₁₀ and (OOG)₁₀ (Figure S7). In particular, a qualitative comparison of the extent of these spectral changes suggests that this stabilizing effect seems to be larger for (OOG)₁₀. This may be related to a possible hydrogen bonding interaction between the Hyp residues in the Xaa and Yaa positions.⁵⁸ These observations, again, confirm a sequence-specific effect of the imposed osmotic pressure changes.

Effect of hydration on the axial length

The sequence-specific molecular response to osmotic pressure is not restricted to the change in the interchain hydrogen bond strength. In the following we show that also the characteristic triple helical structure undergoes significant changes, as witnessed by changes in the helical rise per residue z_p upon hydration. For (PPG)₁₀ a small, gradual decrease in z_p is observed (Figure 3b). In contrast, for (POG)₁₀, z_p first increases slightly until ~50% RH ($\Pi \approx 96$ MPa) and then decreases strongly at higher humidities. Finally, (OOG)₁₀ exhibits a small, gradual increase in z_p until ~50% RH. But then z_p remains constant until the highest RH tested. When measured at full hydration (i.e., when immersed in bulk aqueous medium), the CMPs (PPG)₁₀ and (POG)₁₀ exhibit a further decrease in z_p relative to the 90% RH value. (Figure 3b). This indicates a considerable contraction of the triple helix (up to 2.5%) in the low osmotic pressure range. This results in an overall contraction of (PPG)₁₀ and (POG)₁₀ when moving from the dry to the fully

hydrated state. For (OOG)₁₀, the helical rise per residue could not be determined in bulk water, as this CMP dissolved under these conditions. Overall, also for the helical rise per residue, it is evident that each sequence exhibits its own characteristic response to hydration.

Extrapolating to longer CMPs and natural collagen, a similar overall axial response is expected. On the one hand, the chains are axially staggered by one residue. As a result, maximally 9 out of 10 tripeptide units can fold into a triple helical structure while the peptide termini are disordered. On the other hand, close packing may partially stabilize and align the termini. Based on the available data, no clear conclusion can be drawn about the relative contribution of disordered termini on the obtained XRD pattern. Comparing our results to natural collagen I, the observation that the canonical sequences (PPG)₁₀ and (POG)₁₀ undergo an overall contraction upon hydration (Figure 3b) is unexpected, considering that natural collagen I displays the opposite response and elongates upon hydration. As the canonical collagen tripeptide units PPG and POG are abundant in natural collagen I, this result strongly suggests that other sequences need to elongate in response to hydration to obtain the net elongation of full-length collagen I that has been observed in earlier studies of rat tail collagen and turkey leg tendon.^{25,27}

Effect of hydration on the properties of a host-guest peptide

To address this discrepancy, we consider one example of a host-guest peptide with the guest sequence Ala-Arg-Gly-Ser-Asp-Gly (ARGSDG). This motif, located in the gap region of collagen I (Figure S1), has been predicted to be highly responsive to osmotic pressure.²⁷ To guarantee a stable triple helix configuration, the guest motif was inserted into a PPG host sequence to yield the host-guest CMP (PPG)₄-ARGSDG-(PPG)₄. For this host-guest CMP, the center-to-center distance a is larger than for (PPG)₁₀ (Figure 3c; Figure S8). This larger spacing between triple helices most likely originates from the large size of the Arg side chain that prevents closer packing of the triple helices. Upon hydration, the spacing between triple helices increases in a continuous fashion and with a larger slope than for the PPG host motif (Figure 3c).

Importantly, the host-guest CMP elongates with increasing hydration (Figure 3d) and, therefore, shows the same osmotic pressure response as natural collagen I and as predicted in MD simulations.²⁷ Hence, the guest sequence seems to fundamentally alter the axial behavior of the entire peptide. This is remarkable, as (PPG)₁₀ and the host-guest CMP differ in only four out of thirty amino acids. In fact, considering the one-residue stagger of the individual chains in the triple helix, we assume that the guest sequence affects 3 out of 9 tripeptide units (i.e., ~30% of the overall sequence). Under this assumption, and considering that the host sequence behaves similar to (PPG)₁₀, the net elongation of the host-guest CMP is the result of the superposition of the previously established small contraction of the PPG host sequence and a much larger



elongation of the guest sequence. In a crude approximation, we estimate that the guest sequence contributes, in the range from 5 to 64% RH, with an elongation of 0.14 Å to the overall length change of the host–guest CMP. This means that the elongation of the guest sequence is 10 times larger than the contraction found for the host sequence.

Within the tested RH range, the computed elongation of 0.14 Å corresponds to a relative axial length change of ~5%, a value that roughly compares with the average elongation predicted for the gap region of collagen I using MD simulations (i.e., 10.7%).²⁷ Even though the trend points into the same direction, it needs to be noted that the absolute values of the length change should not be compared. The local structural environment of the guest sequence is different in a PPG host and in natural collagen I. Nevertheless, this result seems to corroborate the hypothesis that specific sequences respond differently to osmotic pressure. It may further indicate that the structural and molecular properties of sequences located in the gap and overlap regions respond differently and are tailored to control the local response of full-length collagen I to hydration. To further corroborate the specific role of sequences in the gap and overlap region, further experiments with a more diverse range of sequences will be needed, possibly considering the fact that collagen I is a heterotrimer.

Discussion and conclusion

This study reveals that osmotic pressure affects CMPs in a sequence-specific manner and that hydration-induced structural changes occur both at the level of individual triple helices and within higher-order assemblies. At the molecular level, increasing hydration causes a strengthening of interchain hydrogen bonding. Individual triple helices further show an osmotic pressure-dependent axial response. CMPs consisting of the abundant tripeptide units PPG and POG contract at high levels of hydration while elongation was observed for the host–guest peptide. Our results thus further support earlier findings that water is an essential part of the collagen molecule.^{13,14,16–19,27,43,62,69} The amino acid sequence determines the local interaction of the triple helix with water and, as a result, the axial response to hydration. Most interestingly, significant axial changes are observed in the near-physiological osmotic pressure range (<1 MPa^{70–72}). Extrapolating to collagen fibers and collagen-based tissues, we conclude that the previously observed hydration-induced response^{25,27} is thus indeed the sum of local and sequence-specific contraction and elongation.

The sequence-specific osmotic pressure response extends to higher-order assemblies. At low hydration levels, the CMPs (PPG)₁₀, (POG)₁₀ and (OOG)₁₀ all show a distorted hexagonal configuration with sequence-specific center-to-center distances between triple helices. Triple helix packing changes into a new CMP characteristic configuration when the triple helical CMPs are immersed in bulk water. This indicates

that the amino acid sequence also determines direct and water-mediated interactions between triple helices. The different packing configurations with a characteristic number of nearest neighbors probably reflects a subtle balance between the strength of direct interhelical bonds and the geometry and strength of water interactions. These findings bear interesting implications for natural systems. They suggest that the local structure and composition of collagen side chains determines the local packing within collagen fibrils. As a consequence, the sequence-specific reorganization of nearest neighbors may play an important regulatory role as the exposure of functional binding sites and channels for ions or mineralization precursors is likely to be altered.^{9,23,26}

Axial responses and the reconfiguration of triple helix packing also affect the mechanical properties of individual triple helices and higher-order assemblies. Hydration-induced contraction or elongation of molecules, fibrils and fibers primarily aids the maintenance of tissue pretension. At the same time, force propagation pathways through collagen assemblies may be altered, possibly even affecting localized bond rupture processes at high strains.⁷³ From a mechanical point of view, collagen has evolved to utilize different structural levels to gather energy from water chemical potential gradients and to efficiently convert it into mechanical energy. Unveiling the molecular basis of this behavior is of utmost importance for understanding the mechanical structure–function relationships of collagen. Molecular level insights will ultimately aid establishing the principles of tissue formation and homeostasis and provide new concepts for the design of bioinspired actuation systems.

Toward such applications, it will be highly interesting to investigate CMPs with a mixed content of PPG, POG and OOG tripeptide units. In particular, the analysis of CMPs with a controlled 3D arrangement and spacing of hydroxyl groups will shed light on possible synergistic effects between neighboring hydroxyl groups. It appears likely that other amino acids with hydrogen bond donating or accepting side chains can substitute for Hyp as an anchoring point for hydrogen bonds. In combination with amino acid size, these properties may eventually be used to control the water structure and thus the osmotic pressure response of CMPs. Experimental and/or simulation studies where individual or several neighboring amino acids are systematically varied will ultimately allow for establishing the relationship between amino acid sequence and lateral and axial osmotic pressure responses.

Materials and methods

Peptides

All CMPs were obtained from commercial suppliers. The canonical CMPs H-(PPG)₁₀-OH and H-(POG)₁₀-OH were purchased from PeptaNova (Sandhausen, Germany). The peptide H-(OOG)₁₀-OH was synthesized by JPT Peptide Technologies GmbH (Berlin, Germany). The host–guest CMP



H-(PPG)₄-ARGSDG-(PPG)₄-OH was received from Centic Biotec (Heidelberg, Germany). The purity of all peptides was >90 percent.

X-ray diffraction

XRD experiments were performed using lyophilized peptides as obtained from the suppliers. The lyophilized CMP powder (~0.1 mg) was placed into an open borosilicate glass capillary. The capillary was connected to a humidity generator (WET-SYS, Caluire, France) to provide air with controlled water content (Figure S2A). During XRD measurements, a flow rate of 70 ml min⁻¹ was used. The temperature and the relative humidity (RH) were monitored by a sensor (Sensirion, Staefa, Switzerland) that was placed close to the sample. RH was converted into Π , according to $\Pi = (RT/V_w) \cdot \ln(\text{RH})$, where R is the gas constant, T the temperature, and V_w the molar volume of water (18 cm³ mol⁻¹).

Experiments were performed at the BESSY II synchrotron (Helmholtz-Zentrum Berlin, μ -Spot beamline). An x-ray beam of 100 μ m in diameter and an energy of 15 keV (wavelength: 0.826 Å) was used. The capillary, connected to the humidity generator, was mounted horizontally on a motorized x , y , z sample stage. To reduce possible beam damage, the sample was moved 100 μ m horizontally between each measurement (1 cm overall displacement perpendicular to the primary beam direction). The scattered intensity was recorded on a 2CCD (MarMosaic 225, Rayonix, Evanston, IL, USA) or an EIGERX (Dectris, Baden, Switzerland) x-ray detector (3072 \times 3072 pixels and 3110 \times 3269 pixels, respectively). The precise position of the beam center, of the sample-to-detector distance (around 310 mm), and of the detector tilt were determined using quartz as a standard. Azimuthal integration of the obtained 2D intensity patterns was performed using the DPDAK (Directly Programmable Data Analysis Kit) software.⁷⁴ The program provides the 1D intensity profiles, $S(q)$, where q denotes the magnitude of the scattering vector. Peaks 1–7 were identified and their q -positions were determined using a home-written Python script. Peak 1 was fitted with a Lorentzian function, while Gaussian fits were used for peaks 2–7. Baseline correction was applied before fitting the respective peaks.

Attenuated total reflection (ATR)—Fourier transform infrared spectroscopy (FTIR)

To perform ATR-FTIR spectroscopy, a film of each CMP was prepared on the ATR crystal. The CMP powder was dissolved in ultrapure water to a concentration of 5 mg ml⁻¹. A drop of this solution (20 μ l) was then spread on the ATR crystal and the water was slowly evaporated under gentle nitrogen flow. The sample was connected to a humidity generator, using a setup similar as for the XRD experiments (Figure S2B). FTIR spectra were collected on a VERTEX 70 FTIR spectrometer (Bruker, Billerica, MA, USA) with 64 scans for each spectrum. The resolution was set to 4 cm⁻¹ and a Blackman-Harris 3-Term apodization function and a Zero-filling factor

of 4 was used. After obtaining a stable spectrum for the film equilibrated in dry conditions (<10% RH), RH was gradually increased up to 80 percent.

Molecular dynamics simulations

The simulation box (Figure S3) was built based on the crystal structure of (OOG)₁₀ (PDB 1wzb).⁵⁹ Starting from this structure, three triplets and all OH groups were removed to obtain (PPG)₇. The CMP triple helices were periodically connected and a total number of 8 \times 8 triple helices were assembled in the simulation box in an antiparallel fashion (Figure S3). An antiparallel triple helix orientation is common in crystal structures of CMPs with free terminal NH₂ and COOH groups.⁴² It allows for neutralizing the local charges in assemblies of tightly packed triple helices. During the simulations, the x , y box angle was fixed at 60° so that the triple helices were forced to maintain a hexagonal lattice structure (Figure S3).

The GROMACS 2016 simulation package⁷⁵ was used. The peptide force field was AMBER 99⁷⁶ in the NpT ensemble, with a semi-isotropic Parrinello-Rahman barostat at 1 bar and a temperature of 300 K, using the velocity rescaling thermostat.⁷⁷ The water model was TIP4p 2005. Periodic boundary conditions were applied to avoid surface effects of triple helices and water. The Smooth Particle-Mesh Ewald method was used for calculating long-range electrostatics. The total simulation time was 50 ns with a 2 fs time step. The trajectories were analyzed using our own Python library (located at <https://gitlab.com/netzlab/maicos>) based on MDAnalysis.⁷⁸

The intensities $S(q_{hkl})$ of the discrete scattering lines for each accessible set of Miller indices (h , k , l) were determined by first “rectangularizing” the triclinic simulation box (Figure S3) and subsequently computing

$$S(q_{hkl}) = \sum_{n,m=1}^N f_n(q)f_m(q)\exp[-iq_{hkl}(r_n - r_m)]$$

for all frames in the trajectory. Here, N is the number of atoms, r_n the position of the n^{th} atom, $q=|q_{hkl}|$ is the magnitude of the scattering vector, and $f_n(q)$ the atomic form factor based on numerical Hartree-Fock wave functions.^{79,80} The line intensities represent the Fourier coefficients of the electron density distribution.

Acknowledgments

We thank the HZB (BESSY II, Helmholtz-Zentrum für Materialien und Energie, Berlin, Germany) for the allocation of synchrotron radiation beam time. We are grateful to Stefan Siegel and Chenghao Li for technical support during x-ray synchrotron experiments at the μ -spot beamline and Klaus Bienert with setting up *in situ* humidity-controlled measurements. This work was supported by the International Max Planck Research School (IMPRS) on Multiscale Bio-Systems (L.R.R.).



Funding

Open Access funding enabled and organized by Projekt DEAL.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Conflict of interest

The authors declare that they have no conflict of interest.

Open access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1557/s43577-021-00138-9>.

References

- 1.P. Fratzl, Ed., *Collagen: Structure and Mechanics* (Springer, New York, 2008)
- 2.S. Ricard-Blum, The collagen family. *Cold Spring Harb. Perspect. Biol.* (2011). <https://doi.org/10.1101/cshperspect.a004978>
- 3.A. Rich, F.H. Crick, The molecular structure of collagen. *J. Mol. Biol.* **3**, 483 (1961)
- 4.W. Traub, A. Yonath, Polymers of tripeptides as collagen models: I. X-ray studies of poly(L-prolyl-glycyl-L-proline) and related polytripeptides. *J. Mol. Biol.* **16**, 404 (1966)
- 5.J.A. Petruska, A.J. Hodge, A subunit model for the tropocollagen macromolecule. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 871 (1964)
- 6.R.D.B. Fraser, T.P. MacRae, A. Miller, E. Suzuki, Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497 (1983)
- 7.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**, 9001 (2006)
- 8.R.D.B. Fraser, T.P. MacRae, A. Miller, Molecular packing in type I collagen fibrils. *J. Mol. Biol.* **193**, 115 (1987)
- 9.Z. Xu, W. Zhao, Z. Wang, Y. Yang, N. Sahai, Structure analysis of collagen fibril at atomic-level resolution and its implications for intra-fibrillar transport in bone biomineralization. *Phys. Chem. Chem. Phys.* **20**, 1513 (2018)
10. G.N. Ramachandran, R. Chandrasekharan, Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers* **6**, 1649 (1968)
11. S. Sakakibara, Y. Kishida, K. Okuyama, N. Tanaka, T. Ashida, M. Kakudo, Single crystals of (Pro-Pro-Gly)₁₀, a synthetic polypeptide model of collagen. *J. Mol. Biol.* **65**, 371 (1972)
12. S. Lees, H.A. Mook, Equatorial diffraction spacing as a function of water content in fully mineralized cow bone determined by neutron diffraction. *Calcif. Tissue Int.* **39**, 291 (1986)
13. J. Bella, M. Eaton, B. Brodsky, H.M. Berman, Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75 (1994)
14. J. Bella, B. Brodsky, H.M. Berman, Hydration structure of a collagen peptide. *Structure* **3**, 893 (1995)

15. G. Melacini, A.M. Bonvin, M. Goodman, R. Boelens, R. Kaptein, Hydration dynamics of the collagen triple helix by NMR. *J. Mol. Biol.* **300**, 1041 (2000)
16. R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, Crystal structure of a collagen-like polypeptide with repeating sequence Pro-Hyp-Gly at 1.4 Å resolution: Implications for collagen hydration. *Biopolymers* **56**, 8 (2001)
17. R.Z. Kramer, J. Bella, B. Brodsky, H.M. Berman, The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *J. Mol. Biol.* **311**, 131 (2001)
18. A. De Simone, L. Vitagliano, R. Berisio, Role of hydration in collagen triple helix stabilization. *Biochem. Biophys. Res. Commun.* **372**, 121 (2008)
19. K.M. Ravikumar, W. Hwang, Region-specific role of water in collagen unwinding and assembly. *Proteins* **72**, 1320 (2008)
20. S. Leikin, V.A. Parsegian, W.-H. Yang, G.E. Walrafen, Raman spectral evidence for hydration forces between collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11312 (1997)
21. S. Leikin, D.C. Rau, V.A. Parsegian, Direct measurement of forces between self-assembled proteins: Temperature-dependent exponential forces between collagen triple helices. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 276 (1994)
22. E.C. Spitzner, S. Röper, M. Zerson, A. Bernstein, R. Magerle, Nanoscale swelling heterogeneities in type I collagen fibrils. *ACS Nano* **9**, 5683 (2015)
23. J. Zhu, C.L. Hoop, D.A. Case, J. Baum, Cryptic binding sites become accessible through surface reconstruction of the type I collagen fibril. *Sci. Rep.* **8**, 16646 (2018)
24. A. Gautieri, S. Vesentini, A. Redaelli, M.J. Buehler, Hierarchical structure and nano-mechanics of collagen microfibrils from the atomistic scale up. *Nano Lett.* **11**, 757 (2011)
25. L. Bertinetti, A. Masic, R. Schuetz, A. Barbetta, B. Seidt, W. Wagermaier, P. Fratzl, Osmotically driven tensile stress in collagen-based mineralized tissues. *J. Mech. Behav. Biomed. Mater.* **52**, 14 (2015)
26. O.G. Andriotis, S. Desissaire, P.J. Thurner, Collagen fibrils: Nature's highly tunable nonlinear springs. *ACS Nano* **12**, 3671 (2018)
27. A. Masic, L. Bertinetti, R. Schuetz, S.W. Chang, T.H. Metzger, M.J. Buehler, P. Fratzl, Osmotic pressure induced tensile forces in tendon collagen. *Nat. Commun.* **6**, 5942 (2015)
28. T. Giesa, R. Schuetz, P. Fratzl, M.J. Buehler, A. Masic, Unraveling the molecular requirements for macroscopic silk supercontraction. *ACS Nano* **11**, 9750 (2017)
29. D. Liu, A. Tarakanova, C.C. Hsu, M. Yu, S. Zheng, L. Yu, J. Liu, Y. He, D.J. Dunstan, M.J. Buehler, Spider dragline silk as torsional actuator driven by humidity. *Sci. Adv.* **5**, eaau9183 (2019)
30. P. Fratzl, R. Elbaum, I. Burgert, Cellulose fibrils direct plant organ movements. *Faraday Discuss.* **139**, 275 (2008)
31. X. Chen, L. Mahadevan, A. Driks, O. Sahin, *Bacillus* spores as building blocks for stimuli-responsive materials and nanogenerators. *Nat. Nanotechnol.* **9**, 137 (2014)
32. J.A.M. Ramshaw, N.K. Shah, B. Brodsky, Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* **122**, 86 (1998)
33. R. Berisio, A.D. Simone, A. Ruggiero, R. Improta, L. Vitagliano, Role of side chains in collagen triple helix stabilization and partner recognition. *J. Pept. Sci.* **15**, 131 (2009)
34. M.D. Shoulders, R.T. Raines, Collagen structure and stability. *Annu. Rev. Biochem.* **78**, 929 (2009)
35. J.A. Fallas, L.E. O'Leary, J.D. Hartgerink, Synthetic collagen mimics: Self-assembly of homotrimers, heterotrimers and higher order structures. *Chem. Soc. Rev.* **39**, 3510 (2010)
36. G.B. Fields, Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.* **8**, 1237 (2010)
37. S. Sakakibara, Y. Kishida, Y. Kikuchi, R. Sakai, K. Kakiuchi, Synthesis of poly-(L-prolyl-L-prolylglycyl) of defined molecular weights. *Bull. Chem. Soc. Jpn.* **41**, 1273 (1968)
38. S. Sakakibara, K. Inouye, K. Shudo, Y. Kishida, Y. Kobayashi, D.J. Prockop, Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta* **303**, 198 (1973)
39. K. Okuyama, N. Tanaka, T. Ashida, M. Kakudo, S. Sakakibara, An x-ray study of the synthetic polypeptide (Pro-Pro-Gly)₁₀. *J. Mol. Biol.* **72**, 571 (1972)
40. R.Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari, H.M. Berman, X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* **280**, 623 (1998)
41. V. Nagarajan, S. Kamitori, K. Okuyama, Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. *J. Biochem.* **125**, 310 (1999)
42. R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)₁₀]₃. *Protein. Sci.* **11**, 262 (2002)
43. K. Okuyama, C. Hongo, R. Fukushima, G. Wu, H. Narita, K. Noguchi, Y. Tanaka, N. Nishino, Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: Implications for proline ring puckering. *Biopolymers* **76**, 367 (2004)
44. K. Okuyama, K. Miyama, M. Mizuno, H.P. Bächinger, Crystal structure of (Gly-Pro-Hyp)₉: Implications for the collagen molecular model. *Biopolymers* **97**, 607 (2012)
45. S. Boudko, S. Frank, R.A. Kammerer, J. Stetefeld, T. Schulthess, R. Landwehr, A. Lustig, H.P. Bächinger, J. Engel, Nucleation and propagation of the collagen triple helix



- in single-chain and trimerized peptides: Transition from third to first order kinetics. *J. Mol. Biol.* **317**, 459 (2002)
46. A.V. Persikov, Y. Xu, B. Brodsky, Equilibrium thermal transitions of collagen model peptides. *Protein Sci.* **13**, 893 (2004)
47. K. Mizuno, S.P. Boudko, J. Engel, H.P. Bächinger, Kinetic hysteresis in collagen folding. *Biophys. J.* **98**, 3004 (2010)
48. Y. Kobayashi, R. Sakai, K. Kakiuchi, T. Isemura, Physicochemical analysis of (Pro-Pro-Gly)_n with defined molecular weight-temperature dependence of molecular weight in aqueous solution. *Biopolymers* **9**, 415 (1970)
49. J. Engel, H.T. Chen, D.J. Prockop, H. Klump, The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)_n and (L-Pro-L-Hyp-Gly)_n. *Biopolymers* **16**, 601 (1977)
50. S.K. Holmgren, K.M. Taylor, L.E. Bretscher, R.T. Raines, Code for collagen's stability deciphered. *Nature* **392**, 666 (1998)
51. K. Kar, P. Amin, M.A. Bryan, A.V. Persikov, A. Mohs, Y.H. Wang, B. Brodsky, Self-association of collagen triple helix peptides into higher order structures. *J. Biol. Chem.* **281**, 33283 (2006)
52. L. Vitagliano, R. Berisio, L. Mazzarella, A. Zagari, Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers* **58**, 459 (2001)
53. C.G. Long, E. Braswell, D. Zhu, J. Apigo, J. Baum, B. Brodsky, Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry* **32**, 11688 (1993)
54. M.S. Ackerman, M. Bhat, N. Shenoy, K. Beck, J.A.M. Ramshaw, B. Brodsky, Sequence dependence of the folding of collagen-like peptides: Single amino acids affect the rate of triple-helix nucleation. *J. Biol. Chem.* **274**, 7668 (1999)
55. A.V. Persikov, J.A. Ramshaw, A. Kirkpatrick, B. Brodsky, Amino acid propensities for the collagen triple-helix. *Biochemistry* **39**, 14960 (2000)
56. J. Emsley, C.G. Knight, R.W. Farnedale, M.J. Barnes, R.C. Liddington, Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* **101**, 47 (2000)
57. R.W. Farnedale, T. Lisman, D. Bihan, S. Hamaia, C.S. Smerling, N. Pugh, A. Konitsiotis, B. Leitinger, P.G. de Groot, G.E. Jarvis, N. Raynal, Cell-collagen interactions: The use of peptide toolkits to investigate collagen-receptor interactions. *Biochem. Soc. Trans.* **36**, 241 (2008)
58. R. Berisio, V. Granata, L. Vitagliano, A. Zagari, Imino acids and collagen triple helix stability: Characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats. *J. Am. Chem. Soc.* **126**, 11402 (2004)
59. 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)-hydroxyprolylglycine]₁₀. *Biochemistry* **44**, 15812 (2005)
60. K. Mizuno, T. Hayashi, D. Peyton, H.P. Bächinger, Hydroxylation-induced stabilization of the collagen triple helix. *J. Biol. Chem.* **279**, 38072 (2004)
61. M. Schumacher, K. Mizuno, H.P. Bächinger, The crystal structure of the collagen-like polypeptide (glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)₉ at 1.55 Å resolution shows up-puckering of the proline ring in the Xaa position. *J. Biol. Chem.* **280**, 20397 (2005)
62. K. Okuyama, C. Hongo, G. Wu, K. Mizuno, K. Noguchi, S. Ebisuzaki, Y. Tanaka, N. Nishino, H.P. Bächinger, High-resolution structures of collagen-like peptides [(Pro-Pro-Gly)₄-Xaa-Yaa-Gly-(Pro-Pro-Gly)₄]: Implications for triple-helix hydration and Hyp(X) puckering. *Biopolymers* **91**, 361 (2009)
63. K. Okuyama, T. Morimoto, H. Narita, T. Kawaguchi, K. Mizuno, H.P. Bächinger, G. Wu, K. Noguchi, Two crystal modifications of (Pro-Pro-Gly)₄-Hyp-Hyp-Gly-(Pro-Pro-Gly)₄ reveal the puckering preference of Hyp(X) in the Hyp(X):Hyp(Y) and Hyp(X):Pro(Y) stacking pairs in collagen helices. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 88 (2010)
64. K. Okuyama, G. Wu, N. Jiravanichanun, C. Hongo, K. Noguchi, Helical twists of collagen model peptides. *Biopolymers* **84**, 421 (2006)
65. N. Kuznetsova, D.C. Rau, V.A. Parsegian, S. Leikin, Solvent hydrogen-bond network in protein self-assembly: Solvation of collagen triple helices in nonaqueous solvents. *Biophys. J.* **72**, 353 (1997)
66. K. Okuyama, Revisiting the molecular structure of collagen. *Connect. Tissue Res.* **49**, 299 (2008)
67. J.W. Brauner, C.R. Flach, R. Mendelsohn, A quantitative reconstruction of the amide I contour in the IR spectra of globular proteins: From structure to spectrum. *J. Am. Chem. Soc.* **127**, 100 (2005)
68. M.A. Bryan, J.W. Brauner, G. Anderle, C.R. Flach, B. Brodsky, R. Mendelsohn, FTIR studies of collagen model peptides: Complementary experimental and simulation approaches to conformation and unfolding. *J. Am. Chem. Soc.* **129**, 7877 (2007)
69. I. Streeter, N.H. de Leeuw, Atomistic modeling of collagen proteins in their fibrillar environment. *J. Phys. Chem. B* **114**, 13263 (2010)
70. A. Maroudas, Biophysical chemistry of cartilaginous tissues with special reference to solute and fluid transport. *Biorheology* **12**, 233 (1975)
71. N.O. Chahine, F.H. Chen, C.T. Hung, G.A. Ateshian, Direct measurement of osmotic pressure of glycosaminoglycan solutions by membrane osmometry at room temperature. *Biophys. J.* **89**, 1543 (2005)
72. E. Han, S.S. Chen, S.M. Klisch, R.L. Sah, Contribution of proteoglycan osmotic swelling pressure to the compressive properties of articular cartilage. *Biophys. J.* **101**, 916 (2011)
73. C. Zapp, A. Obarska-Kosinska, B. Rennekamp, M. Kurth, D.M. Hudson, D. Mercadante, U. Barayeu, T.P. Dick, V. Denysenkova, T. Prisner, M. Bennati, C. Daday, R. Kappl, F. Gräter, Mechanoradicals in tensed tendon collagen as a source of oxidative stress. *Nat. Commun.* **11**, 2315 (2020)
74. G. Benecke, W. Wagermaier, C. Li, M. Schwartzkopf, G. Flucke, R. Hoerth, I. Zizak, M. Burghammer, E. Metwalli, P. Müller-Buschbaum, M. Trebbin, S. Förster, O. Paris, S.V. Roth, P. Fratzl, A customizable software for fast reduction and analysis of large X-ray scattering data sets: Applications of the new DPDAK package to small-angle X-ray scattering and grazing-incidence small-angle X-ray scattering. *J. Appl. Crystallogr.* **47**, 1797 (2014)
75. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19 (2015)
76. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950 (2010)
77. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007)
78. N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319 (2011)
79. D.T. Cromer, J.B. Mann, *X-Ray Scattering Factors Computed from Numerical Hartree-Fock Wave Functions*. (Office of Scientific and Technical Information (OSTI), 1967). <https://doi.org/10.2172/4555098>
80. E. Prince, Ed., *International Tables for Crystallography* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004)