Supporting Information

Sequence-specific response of collagen mimetic peptides to osmotic pressure

Lorena Ruiz-Rodriguez[1,2], Philip Loche[3], Lise Thornfeldt Hansen[3], Roland R. Netz[3], Peter Fratzl[2], Emanuel Schneck[2,4]*, Kerstin G. Blank[1,]*, Luca Bertinetti[2,5]*

[1]*Max Planck Institute of Colloids and Interfaces, Mechano(bio)chemistry, Am Mühlenberg 1, 14476 Potsdam, Germany*

[2]*Max Planck Institute of Colloids and Interfaces, Department of Biomaterials, Am Mühlenberg 1, 14476 Potsdam, Germany*

[3]*Freie Universität Berlin, Department of Physics, Arnimallee 14, 14195 Berlin, Germany*

[4]*Technische Universität Darmstadt, Department of Physics, Hochschulstrasse 8, 64289 Darmstadt, Germany*

[5]*B CUBE - Center for Molecular Bioengineering, Technische Universität Dresden, 01069 Dresden, Germany*

Lorena Ruiz-Rodriguez: jlorena2r@gmail.com; ORCID: 0000-0002-6777-3476
Philip Loche: ploche@physik.fu-berlin.de; ORCID: 0000-0002-9112-0010
Lise Thornfeldt Hansen: lthornfeldt@gmail.com
Roland R. Netz: rnetz@physik.fu-berlin.de; ORCID: 0000-0003-0147-0162
Peter Fratzl: Peter.Fratzl@mpikg.mpg.de; ORCID: 0000-0003-4437-7830
Emanuel Schneck*: schneck@fkp.tu-darmstadt.de; ORCID: 0000-0001-9769-2194
Kerstin G. Blank*: kerstin.blank@mpikg.mpg.de; ORCID: 0000-0001-5410-6984
Luca Bertinetti*: luca.bertinetti@tu-dresden.de; ORCID: 0000-0002-4666-9610
*corresponding authors

TABLE OF CONTENTS

Estimation of the water content in hexagonally packed triple helices

To estimate the hydration level in the XRD experiments, it is necessary to first know the molecular volumes of one tripeptide repeat unit $\bar{V}_t$ and then that of the water molecules $\bar{V}_w$ around the triple helices. In the following, we describe the general procedure used to estimate the number of water molecules per amino acid (N_W/N_{AA}) for hexagonally packed CMPs, using $(PPG)_n$ as an example. The same procedure can be used to estimate N_W/N_{AA} for $(POG)_n$ and $(OOG)_n$.

The molar volume of a PPG tripeptide unit in its triple helical state is 169.4 $cm^3\ mol^{-1}$,[1] which corresponds to $\bar{V}_t$ = 281.3 $Å^3$. Assuming that PPG is a cylinder with an effective radius r_{eff}, and considering that $(PPG)_n$ adopts a 7/2 structure, the diameter of the helix can be written as

$$r_{eff} = \sqrt{\frac{V}{\pi h}}$$

where V is the volume of the cylinder and h is its height. For a triple helix with a 7/2 structure, 7 triplets correspond to a height of ~20 Å. With $V = 7\bar{V}_t$, the effective diameter equals $2r_{eff}$ = 1.12 nm.

$\bar{V}_w$ was estimated from crystallographic data (pdb 1a3j[2]) as

$$\bar{V}_w = \frac{(V_{cell} - n_h \cdot n_t \cdot \bar{V}_t)}{n_w}$$

where n_h is the number of triple helices and n_t is the number of tripeptide units per helix. V_{cell} is the volume of the unit cell and n_w is the number of water molecules in the unit cell. For example, considering a unit cell as in the pdb structure 1a3j[2] with n_h = 4, n_t = 7, and n_w = 160, $\bar{V}_w$ is estimated to be ~42 $Å^3$, which is significantly larger than for bulk liquid water (~30 $Å^3$). Also, according to the pdb entry, the number of AA in the unit cell is 84 (i.e. $3 \cdot n_h \cdot n_t$). This ultimately yields N_w/N_{AA} = 1.9. Similarly, for (PPG)10 as described in Berisio et al., N_w/N_{AA} = 2.0. Also in the case of POG (Okuyama 2004), there are 42 water molecules for 7 triplets, that translates to 2.0 N_w/N_{AA}.

In the next step, we used $\bar{V}_w$ and r_{eff} to assess the hydration level (N_w/N_{AA}) of the CMPs in a hexagonal arrangement. In the geometry shown in Figure 1, the red triangular element contains 0.5 helices. As the volume of a helix is $V = \pi h r_{eff}^2$, and considering again a 7/2 structure, N_w/N_{AA} can be calculated as

$$\frac{N_w}{N_{AA}} = \bar{h}\frac{\left(\frac{\sqrt{3}}{2}a^2 - \pi r_{eff}^2\right)}{\bar{V}_w}$$

where $\bar{h}$ is the height of the triple helix divided by the number of amino acid therein contained (for a 7/2 structure, $\bar{h}$ is 20/(7·3) Å/AA). For a = 1.28 nm, i.e. the center-to-center distance at the highest experimentally obtained *R.H.*, $N_w/N_{AA} \approx 1.1$. At N_w/N_{AA} = 0.8 and N_w/N_{AA} = 0.4, the center-to-center distance is 1.23 nm and 1.15 nm. This corresponds to an intermediate *R.H.* of ~60 % and very dry conditions, respectively. It must be noted that the considered volumetric relationships could be very different at low hydrations as the partial molar volume of water may change significantly in this hydration range. Therefore, the reported calculations are not exact.

They are only used here as a tool to roughly estimate the water content in hexagonally packed helices.

α1 GPMGPSGPRGLOGPOGAOGPQGFQGPOGEOGEOGASGPMGPRGPOGPOGK
α1 GPMGPSGPRGLOGPOGAOGPQGFQGPOGEOGEOGASGPMGPRGPOGPOG
α2 GPMGLMGPRGPOGAAGAOGPQGFQGPAGEOGEOGQTGPAGARGPAGPO

α1 NGDDGEAGKPGROGERGPOGPQGARGLOGTAGLOGMKGHRGFSGLDGAKG
α1 KNGDDGEAGKPGROGERGPOGPQGARGLOGTAGLOGMKGHRGFSGLDGAK
α2 GKAGEDGHOGKPGROGERGVVGPQGARGFOGTOGLOGFKGIRGHNGLDGL

α1 DAGPAGPKGEOGSOGENGAOGQMGPRGLOGERGROGAOGPAGARGNDGAT
α1 GDAGPAGPKGEOGSOGENGAOGQMGPRGLOGERGROGAOGPAGARGNDGA
α2 KGQOGAPGVKGEOGAOGENGTOGQTGARGLOGERGRVGAOGPAGARGSDG

α1 GAAGPOGPTGPAGPOGFOGAVGAKGEAGPQGPRGSEGPQGVRGEOGPOGP
α1 TGAAGPOGPTGPAGPOGFOGAVGAKGEAGPQGPRGSEGPQGVRGEOGPOG
α2 SVGPVGPAGPIGSAGPOGFOGAOGPKGEIGAVGNAGPAGPAGPRGEVGLO

α1 AGAAGPAGNOGADGQOGAKGANGAOGIAGAOGFOGARGPSGPQGPGGPOG
α1 PAGAAGPAGNOGADGQOGAKGANGAOGIAGAOGFOGARGPSGPQGPGGPO
α2 GLSGPVGPOGNPGANGLTGAKGAAGLOGVAGAOGLOGPRGIOGPVGAAGA

α1 PKGNSGEOGAOGSKGDTGAKGEOGPVGVQGPOGPAGEEGKRGARGEOGPT
α1 GPKGNSGEOGAOGSKGDTGAKGEOGPVGVQGPOGPAGEEGKRGARGEOGP
α2 TGARGLVGEOGPAGSKGESGNKGEOGSAGPQGPOGPSGEEGKRGPNGEAG

α1 GLOGPOGERGGOGSRGFOGADGVAGPKGPAGERGSOGPAGPKGSOGEAGR
α1 TGLOGPOGERGGOGSRGFOGADGVAGPKGPAGERGSOGPAGPKGSOGEAG
α2 SAGPPGPOGLRGSOGSRGLOGADGRAGVMGPPGSRGASGPAGVRGPNGDA

α1 OGEAGLOGAKGLTGSOGSOGPDGKTGPOGPAGQDGROGPPGPOGARGQAG
α1 ROGEAGLOGAKGLTGSOGSOGPDGKTGPOGPAGQDGROGPPGPOGARGQA
α2 GROGEOGLMGPRGLOGSOGNIGPAGKEGPVGLOGIDGROGPIGPAGARGE

α1 VMGFOGPKGAAGEOGKAGERGVOGPOGAVGPAGKDGEAGAQGPOGPAGPA
α1 GVMGFOGPKGAAGEOGKAGERGVOGPOGAVGPAGKDGEAGAQGPOGPAGP
α2 PGNIGFOGPKGPTGDOGKNGDKGHAGLAGARGAOGPDGNNGAQGPOGPQG

α1 GERGEQGPAGSOGFQGLOGPAGPOGEAGKOGEQGVOGDLGAOGPSGARGE
α1 AGERGEQGPAGSOGFQGLOGPAGPOGEAGKOGEQGVOGDLGAOGPSGARG
α2 VQGGKGEQGPPGPOGFQGLOGPSGPAGEVGKOGERGLHGEFGLOGPAGPR

α1 RGFOGERGVQGPOGPAGPRGANGAOGNDGAKGDAGAOGAOGSQGAOGLQG
α1 ERGFOGERGVQGPOGPAGPRGANGAOGNDGAKGDAGAOGAOGSQGAOGLQ
α2 GERGPPGESGAAGPTGPIGSRGPSGPOGPDGNKGEPGVVGAVGTAGPSGP

α1 MOGERGAAGLOGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPOGPAGAO
α1 GMOGERGAAGLOGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPOGPAGA
α2 SGLOGERGAAGIOGGKGEKGEPGLRGEIGNPGRDGARGAOGAVGAOGPAG

α1 GDKGESGPSGPAGPTGARGAOGDRGEPGPOGPAGFAGPOGADGQOGAKGE
α1 OGDKGESGPSGPAGPTGARGAOGDRGEPGPOGPAGFAGPOGADGQOGAKG
α2 ATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQOGAK

α1 OGDAGAKGDAGPOGPAGPAGPOGPIGNVGAOGAKGARGSAGPOGATGFOG
α1 EOGDAGAKGDAGPOGPAGPAGPOGPIGNVGAOGAKGARGSAGPOGATGFO
α2 GERGAKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPGMTGF

α1 AAGRVGPOGPSGNAGPOGPOGPAGKEGGKGPRGETGPAGROGEVGPOGPO
α1 GAAGRVGPOGPSGNAGPOGPOGPAGKEGGKGPRGETGPAGROGEVGPOGP
α2 OGAAGRTGPOGPSGISGPOGPOGPAGKEGLRGPRGDQGPVGRTGEVGAVG

α1 GPAGEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLOGP
α1 OGPAGEKGSOGADGPAGAOGTPGPQGIAGQRGVVGLOGQRGERGFOGLOG
α2 PPGFAGEKGPSGEAGTAGPOGTPGPQGLLGAOGILGLOGSRGERGLOGVA

α1 SGEOGKQGPSGASGERGPOGPMGPOGLAGPOGESGREGAOGAEGSOGRDG
α1 PSGEOGKQGPSGASGERGPOGPMGPOGLAGPOGESGREGAOGAEGSOGRD
α2 GAVGEOGPLGIAGPOGARGPOGAVGSOGVNGAOGEAGRDGNOGNDGPOGR

α1 SOGAKGDRGETGPAGPOGAOGAOGAPGPVGPAGKSGDRGETGPAGPTGPV
α1 GSOGAKGDRGETGPAGPOGAOGAOGAPGPVGPAGKSGDRGETGPAGPTGP
α2 DGQOGHKGERGYOGNIGPVGAAGAPGPHGPVGPAGKHGNRGETGPSGPVG

α1 GPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPOGPOGSOGE
α1 VGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPOGPOGSOG
α2 PAGAVGPRGPSGPQGIRGDKGEPGEKGPRGLPGLKGHNGLQGLPGIAGHH

α1 QGPSGASGPAGPRGPOGSAGAOGKDGLNGLOGPIGOOGPRGRTGDAGPVG
α1 EQGPSGASGPAGPRGPOGSAGAOGKDGLNGLOGPIGOOGPRGRTGDAGPV
α2 GDQGAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVGPAGIRGPQGHQGP

α1 POGPOGPOGPOGPP
α1 GPOGPOGPOGPOGPP
α2 AGPOGPOGPOGPPGVS

Figure S1. Amino acid sequence of human collagen I (α1 chain: Swissprot P02452; α2 chain: Swissprot P08123), highlighting the staggered arrangement of the α1 and α2 chains.[3] The prodomains and telopeptides are not shown. The overlap regions[4] are shaded in grey. The hydroxyproline (O) posttranslational modification is assigned according to a published X-ray structure of rat collagen I (PDB 3hr2).[5] The PPG and POG tripeptide units are highlighted. The OOG tripeptide unit located in the fifth overlap region contains 3*R*-hydroxy-2*S*-proline in the Xaa position[6] so that this OOG unit is not identical to the OOG tripeptide unit studied here. Red sequences have been predicted to be responsive to changes in osmotic pressure.[7] The sequence GARGSD (α2 chain) is inserted into the host-guest peptide used in this study.

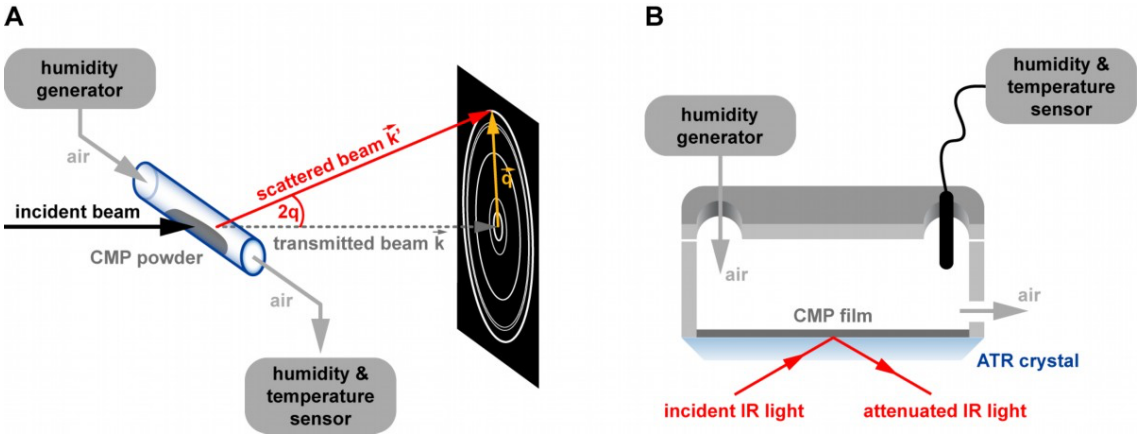

Figure S2. Experimental setups for the *in situ* control of the relative humidity. A) Humidity control during X-ray scattering experiments. The lyophilized CMP powder was placed inside a glass capillary. Air of controlled relative humidity (*R.H.*), provided by a humidity generator, was flushed through the capillary. The *R.H.* values reported throughout the manuscript were obtained from a humidity and temperature measurement at the exit of the capillary. B) FTIR in a humidity-controlled environment. Dried CMP films were prepared on the ATR crystal from aqueous CMP solutions. The ATR crystal was subsequently placed into a chamber and air of controlled *R.H.* was flown through the chamber. The chamber was equipped with a humidity and temperature sensor.

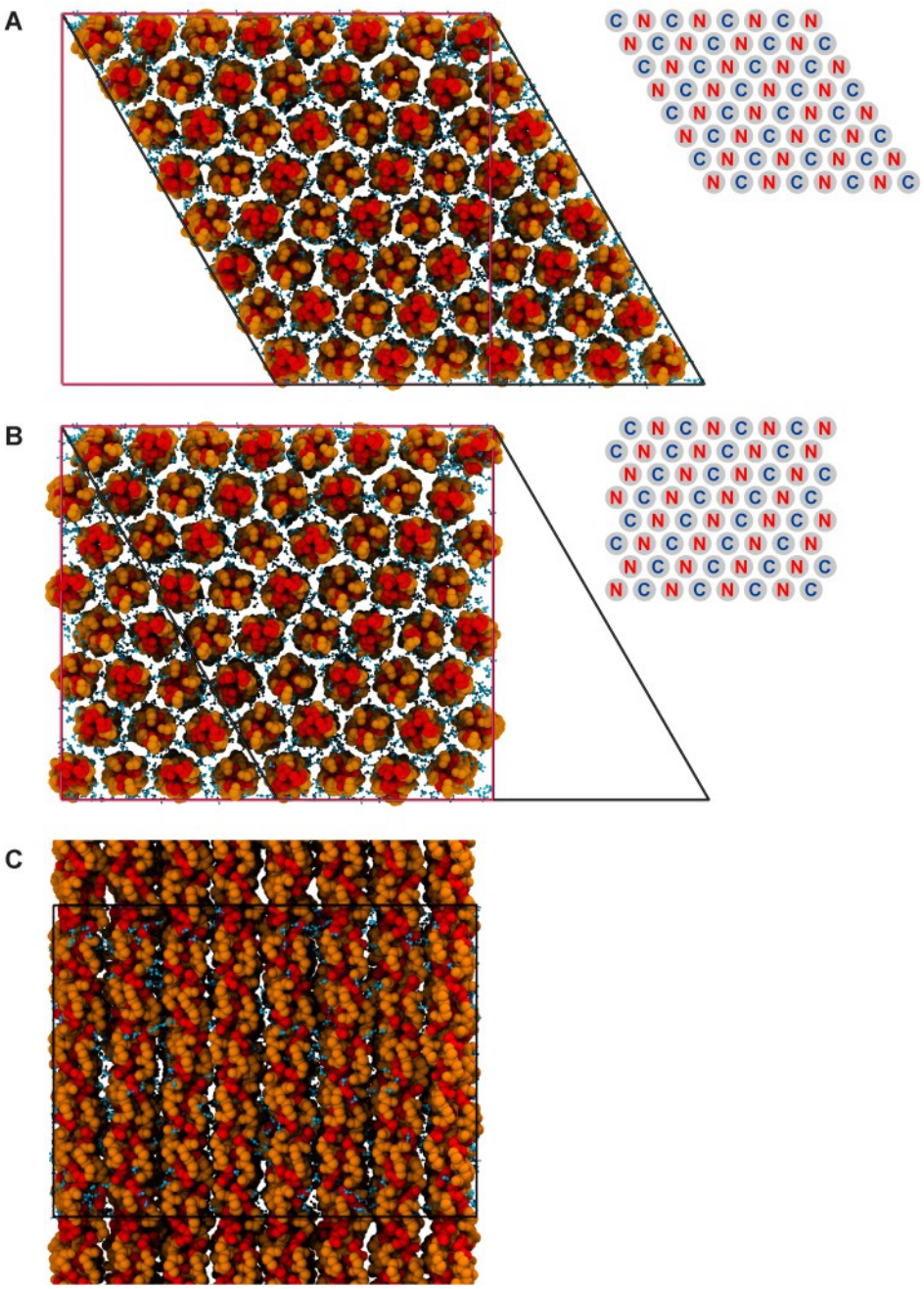

Figure S3. Setup for the molecular dynamics simulations. (A) Top view showing the hexagonal organization of PPG-based triple helices under dehydrated conditions (0.4 N_W/N_{AA}). The simulations were performed with 8×8 quasi-infinitely-long triple helices in an antiparallel arrangement (see N →C orientations in the miniaturized schematic illustration). Each triple helix contains 7 PPG repeats and is connected on both ends with its own periodic image. The simulation box is highlighted as a black trapezoid. The box dimensions were 9.7 nm × 9.7 nm × 6.1 nm while the angle between the two base vectors was 120 degrees. These parameters were fixed during the entire simulation time. Periodic boundary conditions were applied in all 3 dimensions. (B) Simulation box (red rectangle) after "rectangularization", required for the calculation of the structure factor $S(q)$. (C) Side view of the simulation box (black rectangle) showing the quasi-infinite length of the triple helices in the axial direction.

Table S1. Comparison of $(PPG)_{10}$ structural parameters obtained from XRD and MD at dry conditions (XRD: *R.H.* ≈ 10 %; MD: 0.4 N_W/N_{AA}). For all peaks n, the experimental values of q_{n_XRD} and d_{n_XRD} are given. For the simulated data, q_{n_MD} as well as the Miller indices are provided. The Miller indices (h_{MD}, k_{MD} and l_{MD}) refer to the unit cell containing 8×8 CMPs, as used in the MD simulations. For example, a (0,8,0) plane in the simulations corresponds to a (0,1,0) plane in a standard hexagonal cell.

Peak n	d_{n_XRD} **(nm)**	q_{n_XRD} **(nm^{-1})**	q_{n_MD} **(nm^{-1})**	h_{MD}	k_{MD}	l_{MD}
1	1.002	6.30	6.03	0	8	0
2	0.704	8.92	8.50	0	8	6
3	0.553	11.36	10.39	16	0	0
4	0.500	12.56	12.01	0	16	0
5	0.484	12.98	12.07	8	12	6
6	0.433	14.50	13.48	16	8	6
7	0.285	22.04	21.41	0	0	21

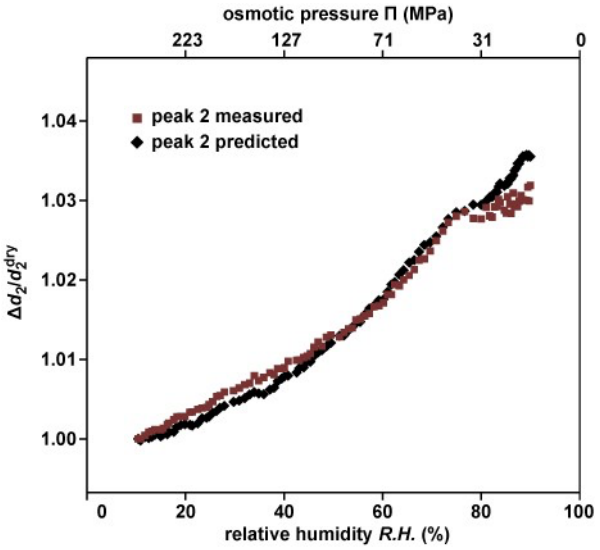

Figure S4. Validation of the lattice plane assignment. Assuming that the lattice planes predicted by MD are correct, the changes in the Bragg distances can be obtained from the experimental data. For example, considering the angular relationships suggested in Table 1, it is predicted from the simulations that peak 2 is related to a lattice plane with an in-plane component (0,8,0) and an axial component (0,0,6). The angle between this lattice plane and the x,y plane, at any relative humidity, is described by the relationship $\alpha^{RH} = \tan^{-1}\left(\frac{21\, d_7^{RH}}{6\, d_1^{RH}}\right)$, where d_1^{RH} *and* d_7^{RH} are the experimental values of d_1 and d_7 at any value of *R.H.*. If this assumption is correct, $d_2^{RH} = d_1^{RH} \sin \alpha^{RH}$ at any *R.H.*. The fact that the calculated $\frac{\Delta d_2}{d_2^{dry}} = \frac{d_2^{RH} - d_2^{10\%}}{d_2^{10\%}}$, where $d_2^{10\%}$ refers to the position of d_2 at the lowest experimentally used *R.H.*, falls onto the experimental curve for d_2 confirms that the CMP can be correctly described with the packing of the MD simulations.

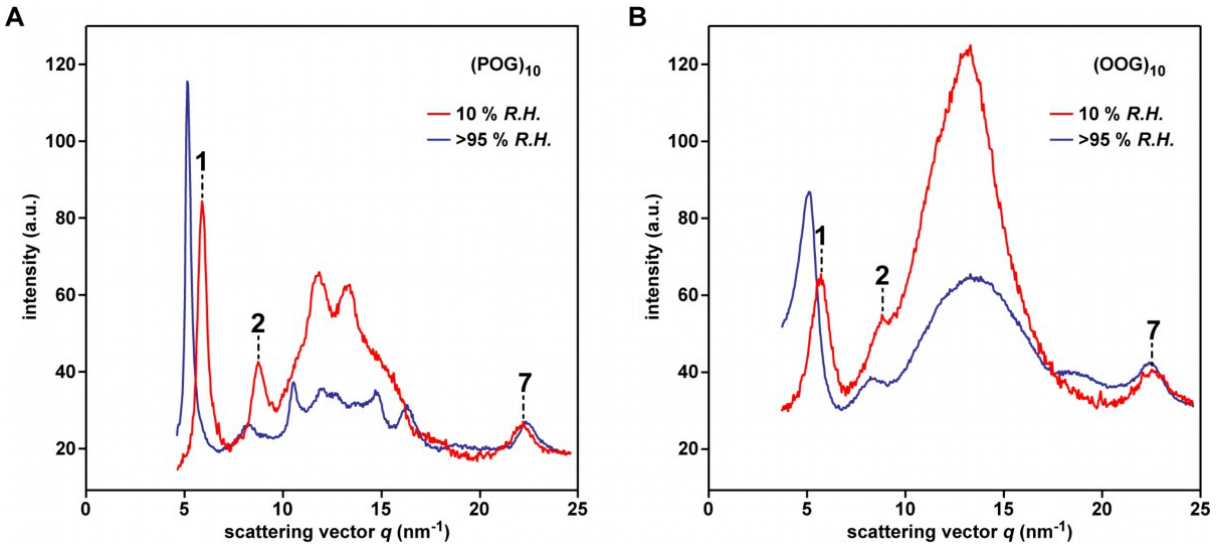

Figure S5. Osmotic pressure response of $(POG)_{10}$ and $(OOG)_{10}$. (A) Scattering profile of $(POG)_{10}$, measured at the lowest (10 %) and highest accessible (>95 %) relative humidity (*R.H.*). (B) Scattering profile of $(OOG)_{10}$, measured at the lowest (10 %) and highest accessible (>95 %) *R.H.*.

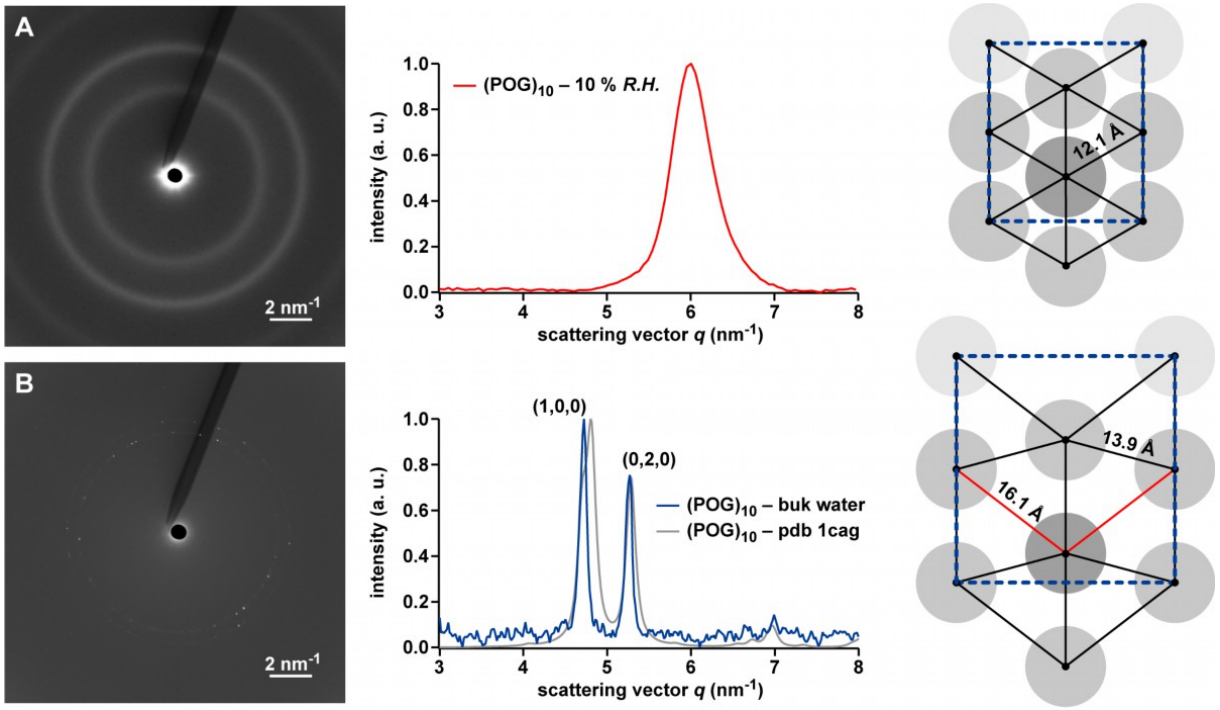

Figure S6. Comparison of the lateral packing arrangement of $(POG)_{10}$ in the dehydrated and fully hydrated state. A) Lateral packing arrangement in dry conditions (10 % *R.H.*). The 2D scatter pattern and the $S(q)$ profile highlight peak 1 that was assigned to the lateral center-to-center distance between triple helices. B) Lateral packing arrangement of triple helices in bulk water (5 mg ml^{-1}). The 2D scatter pattern and the $S(q)$ profile shows the appearance of two new peaks. These peaks are consistent with the diffraction pattern calculated for the rotationally averaged crystal structure PDB 1cag.[8] In this configuration, the number of nearest neighbors is four. The distances to the first (nearest) neighbors are shown as black lines, while the second neighbors are highlighted in red.

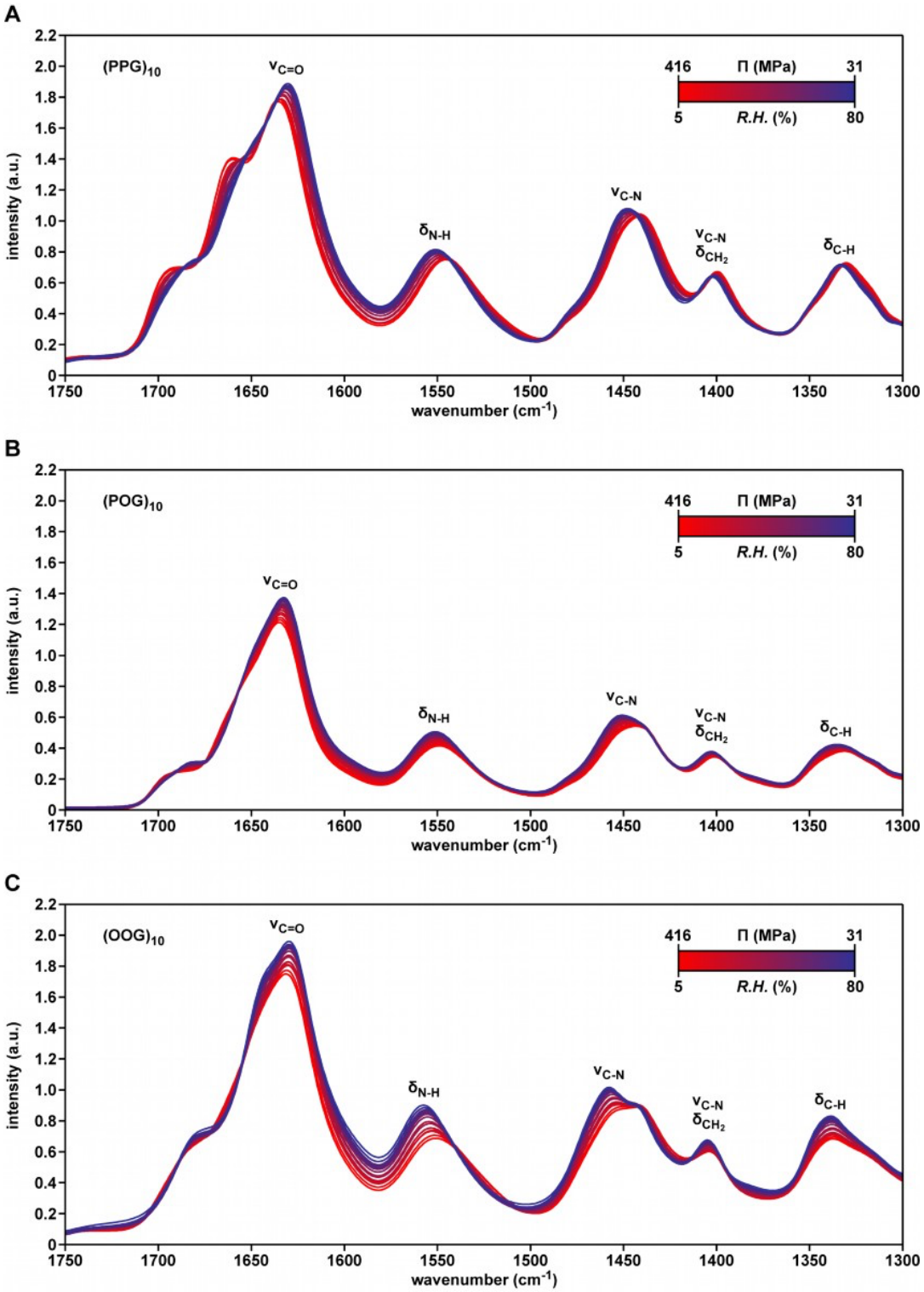

Figure S7. Osmotic pressure effects on the interchain hydrogen bond, monitored by FTIR. A) FTIR spectra of $(PPG)_{10}$, recorded when increasing the relative humidity (*R.H.*) from 5 % to 80 %. B) FTIR spectra of $(POG)_{10}$. C) FTIR spectra of $(OOG)_{10}$. The spectra shown correspond to steps of ~5 % *R.H.*.

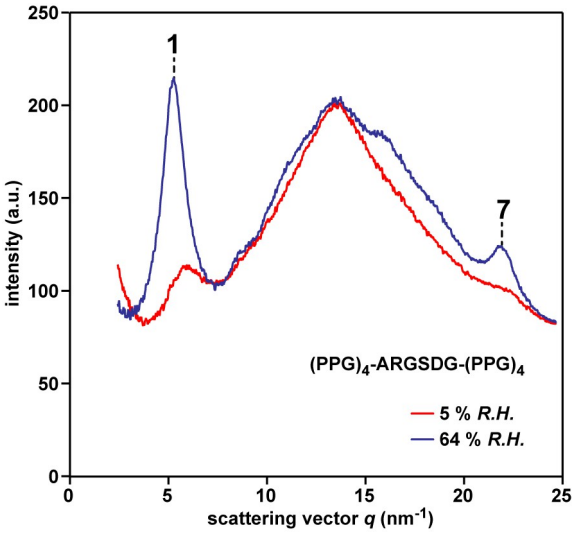

Figure S8. Osmotic pressure response of $(PPG)_4$-ARGSDG-$(PPG)_4$. The scattering profiles at the lowest (5 %) and highest (64 %) measured *R.H.* values are shown. Assuming that the PPG host sequence and the region of the guest sequence respond as independent units, one may expect two components under peak 7. As the expected shift is small (~1 nm^{-1}) compared to the peak width (FWHM ~1 nm^{-1}) it is not possible to resolve these different components.

References

1. 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).
2. 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).
3. A.A. Jalan, D. Sammon, J.D. Hartgerink, P. Brear, K. Stott, S.W. Hamaia, E.J. Hunter, D.R. Walker, B. Leitinger, R.W. Farndale, Chain alignment of collagen I deciphered using computationally designed heterotrimers. *Nat. Chem. Biol.* **16**, 423 (2020).
4. R.D.B. Fraser, T.P. MacRae, A. Miller, E. Suzuki, Molecular conformation and packing in collagen fibrils. *J. Mol. Biol.* **167**, 497 (1983).
5. 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).
6. M.A. Weis, D.M. Hudson, L. Kim, M. Scott, J.-J. Wu, D.R. Eyre, Location of 3-Hydroxyproline Residues in Collagen Types I, II, III, and V/XI Implies a Role in Fibril Supramolecular Assembly. *J. Biol. Chem.* **285**, 2580 (2010).
7. 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).
8. 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 A resolution: implications for proline ring puckering. *Biopolymers* **76**, 367 (2004).