SUPPORTING INFORMATION

The effect of protein backbone hydration on the amide vibrations in Raman and Raman optical activity spectra

Carl Mensch, Patrick Bultinck and Christian Johannessen

CONTENTS

1 Bond lengths (C=O, C-N and N-H)

For the geometry optimised molecular structures depicted in Figure 3, the C=O, C-N and N-H bond lengths are given in Table S 1, Table S 2 and Table S 3 respectively. In this particular hydration configuration, only carbonyl (C=O)^{3} does not from a hydrogen bond with an explicit water molecule and of the N-H groups, only (N-H)^{6} forms a hydrogen bond with an explicit water molecule (see labelling in Fig. S 1).

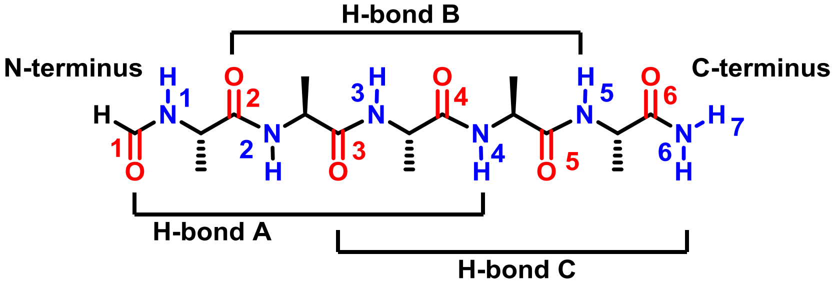


Fig. S 1 Hydrogen bonding patterns in α-helical structure and C=O and N-H labelling.

Table S 1: Difference ΔBL (Å) of the carbonyl (C=O) bond lengths for α-helical structure b-f (see Figure 3) with bond length BL in the gas phase (a) for each residue. $\Delta BL = BL_x - BL_a$ with x = b, c, d, e, f.

			(C=O)^1	(C=O)^2	(C=O)^3	(C=O)^4	(C=O)^5	(C=O)^6
BL	*a*	*gas phase*	*1.2224*	*1.2222*	*1.2226*	*1.2219*	*1.2207*	*1.2217*
ΔBL	b	only-PCM	0.0094	0.0096	0.0094	0.0083	0.0101	0.0091
ΔBL	e	C=O...waters	0.0154	0.0171	0.0086	0.0196	0.0164	0.0123
ΔBL	f	N-H...waters	0.0109	0.0103	0.0095	0.0085	0.0101	0.0092
ΔBL	d	CO and NH waters	0.0176	0.0176	0.0088	0.0197	0.0165	0.0125
ΔBL	c	3 Å shell	0.0173	0.0200	0.0079	0.0216	0.0180	0.0131

Table S 2: Difference ΔBL (Å) of the amide C-N bond lengths for α-helical structure b-f (see Figure 3) with bond length BL in the gas phase (a) for each residue. $\Delta BL = BL_x - BL_a$ with x = b, c, d, e, f.

			(C-N)^1	(C-N)^2	(C-N)^3	(C-N)^4	(C-N)^5	(C-N)^6
BL	*a*	*gas phase*	*1.3563*	*1.3579*	*1.3568*	*1.3542*	*1.3584*	*1.3537*
ΔBL	b	only C-PCM	-0.0087	-0.0064	-0.0064	-0.0038	-0.0068	-0.0073
ΔBL	e	C=O...waters	-0.0144	-0.0131	-0.0055	-0.0162	-0.0159	-0.0174
ΔBL	f	N-H...water	-0.0094	-0.0065	-0.0066	-0.0038	-0.0066	-0.0074
ΔBL	d	CO and NH waters	-0.0154	-0.0133	-0.0053	-0.0161	-0.0149	-0.0175
ΔBL	c	3 Å shell	-0.0174	-0.0136	-0.0039	-0.0149	-0.0145	-0.0214

Table S 3: Difference ΔBL (Å) of the amide N-H bond lengths for α-helical structure b-f (see Figure 3) with bond length BL in the gas phase (a) for each residue. ΔBL = $BL_x - BL_a$ with x = b, c, d, e, f.

			(N-H)^1	**(N-H)^2**	**(N-H)^3**	**(N-H)^4**	**(N-H)^5**	**(N-H)^6**	**(N-H)^7**
BL	*a*	*gas phase*	*1.0097*	*1.0095*	*1.0108*	*1.0128*	*1.0131*	*1.0099*	*1.0079*
ΔBL	b	only C-PCM	0.0015	0.0008	-0.0013	0.0021	0.0027	0.0046	-0.0003
ΔBL	e	C=O...waters	0.0021	0.0017	-0.0011	0.0056	0.0046	0.0047	0.0053
ΔBL	f	N-H...water	0.0121	0.0018	-0.0009	0.0024	0.0028	0.0048	-0.0003
ΔBL	d	CO and NH waters	0.0122	0.0033	-0.0010	0.0057	0.0047	0.0047	0.0060
ΔBL	c	3 Å shell	0.0078	0.0060	0.0021	0.0061	0.0052	0.0053	0.0032

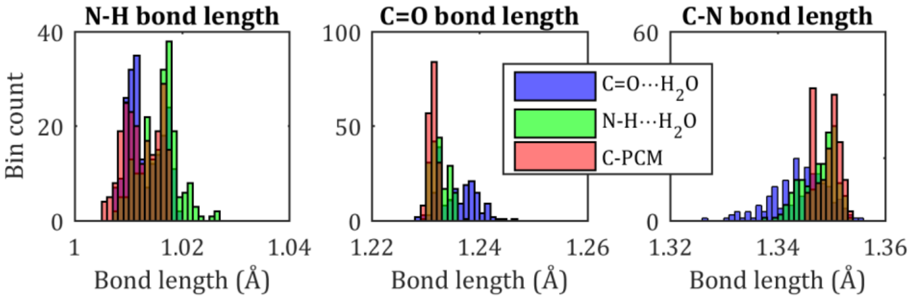
Table S 4: Number of times a N-H group has formed a hydrogen bond with the explicit water molecules (see Figure 7 and Figure 8) in one of the 30 water configurations for each peptide conformation (so maximum is 30). For example, in 24 out of the 30 hydration configurations of α-helix (-64°;-47°), the first $(N\text{-}H)_1$ group has formed a hydrogen bond with one of the water molecules.

	(N-H)^1	**(N-H)^2**	**(N-H)^3**	**(N-H)^4**	**(N-H)^5**	**(N-H)^6**	**(N-H)^7**
α-helix (-64°;-47°)	24	20	26	0	0	0	23
α-helix (-59°;-44°)	28	22	24	0	0	0	24
α-helix (-77°;-34°)	25	21	19	0	0	0	24
α-helix (-66°;-41°)	24	26	19	0	0	0	23
3_{10}-helix (-71°;-18°)	26	14	1	0	0	1	28
LH α-helix (66°;41 °)	24	19	19	0	0	0	26
PPII (-75°;145°)	25	26	28	27	23	28	25
β-strand (-125°,150°)	30	28	25	29	30	28	26

Table S 5: Number of times a C=O group has formed a hydrogen bond with the explicit water molecules (see Figure 7 and Figure 8) in one of the 30 water configurations for each peptide conformation.

	(C=O)^1	**(C=O)^2**	**(C=O)^3**	**(C=O)^4**	**(C=O)^5**	**(C=O)^6**
α-helix (-64°;-47°)	9	6	18	29	27	29
α-helix (-59°;-44°)	19	21	23	27	27	29
α-helix (-77°;-34°)	14	23	25	28	30	30
α-helix (-66°;-41°)	10	18	20	28	27	30
3_{10}-helix (-71°;-18°)	27	21	27	30	29	30
LH α-helix (66°;41 °)	22	12	21	26	27	30
PPII (-75°;145°)	28	29	30	29	30	30
β-strand (-125°,150°)	30	28	27	30	27	29

(a) All amides



(b) Only H_2O bonded and excluding amide 1 and 6

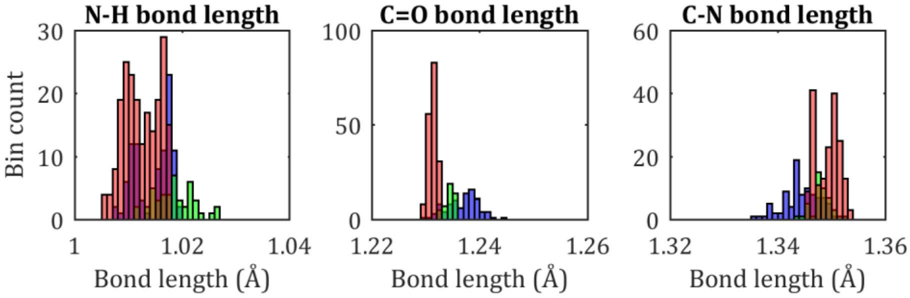
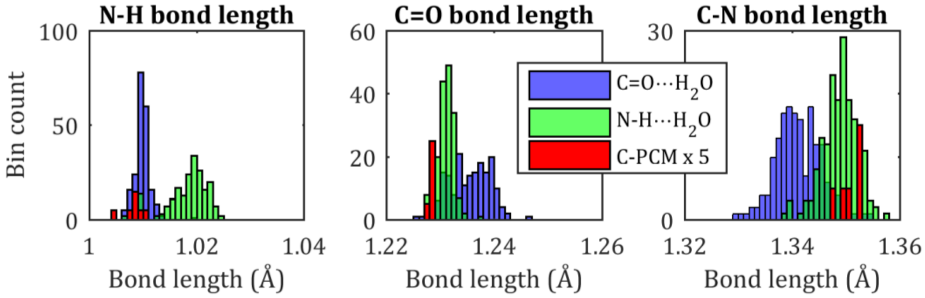


Fig. S 2 Bond length analysis of the amide group (N-H, C=O, C-N) of the hydration of α-helical structure (-66°;-41°). The histograms show the bond lengths of (blue) the ensemble which only included water molecules hydrogen bonded to the C=O groups, compared to (green) when only the water molecules hydrogen bonded to the N-H group were included or (red) when only C-PCM was used without explicit water molecules. The latter ensemble was created by removing all explicit water molecules from the ensemble with the 3 Å explicit hydration shell and rerunning the calculations for all 30 structures in the ensemble with only C-PCM. The bin widths are 0.001 Å.

(a) All amides



(b) Only H_2O bonded and excluding amide 1 and 6

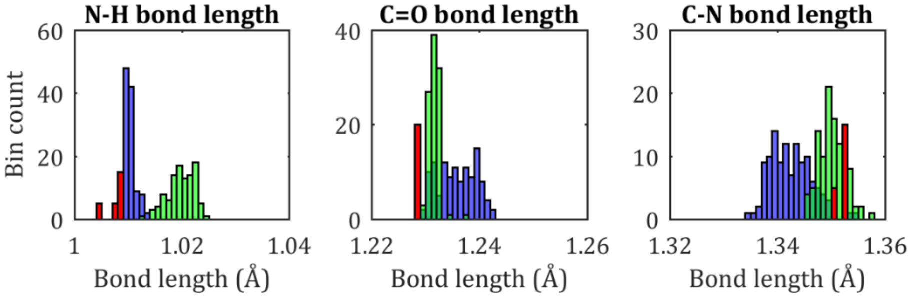


Fig. S 3 Bond length analysis of the amide group (N-H, C=O, C-N) of the hydration of PPII structure (blue). The bond lengths of the ensemble which only included water molecules hydrogen bonded to the C=O groups, compared to (green) when only the water molecules hydrogen bonded to the N-H group were included or (red) the PPII reference structure without explicit water molecules (a single structure). In all calculations C-PCM was used. The red bars were multiplied by 5 for better visual representation. Bin widths are 0.001 Å.

2 Explicit hydration with gas phase or C-PCM

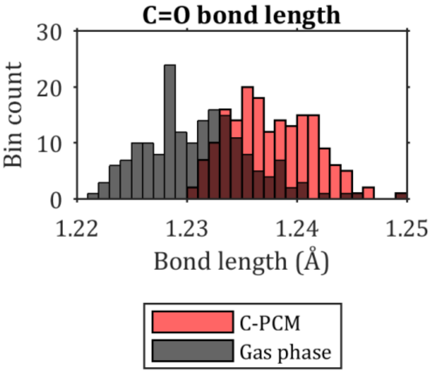


Fig. S 4 Histograms of all C=O bond lengths in 30 solvent configurations with the mean backbone angles of the solute close to ϕ;ψ ≈ -66°;-41°. (red) C-PCM is included in the partial geometry optimization to represent the bulk solvent, (black) only explicit water molecules are included to represent the solvent and the entire system is calculated in the gas phase. Bin widths are 0.001 Å.

3 Basis set dependence

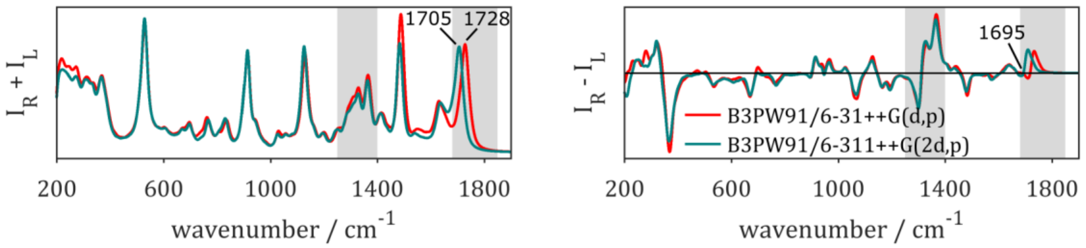


Fig. S 5 Effect of the basis set on the Raman (I_R+I_L) and ROA (I_R-I_L) spectra averaged over 30 water configurations of HCO-(L-Ala)$_5$-NH$_2$ in an α-helical conformation (-66°;-41°). The basis set is either 6-31G(d,p) or 6-311G(2d,p) in the geometry optimization and the respective basis set augmented with diffuse functions (++) on all atoms for the Raman and ROA calculation.

4 Similarity analysis

The similarity S_{fg} of calculated spectrum f and an experimental spectrum g in the entire 300-1800 cm^{-1} spectral window is calculated using the following overlap integral,

$$S_{fg} = \frac{\int_{\tilde{\nu}\geq 300}^{\tilde{\nu}\leq 1800} f(\sigma\tilde{\nu})g(\tilde{\nu})d\tilde{\nu}}{\sqrt{\int_{\tilde{\nu}\geq 300}^{\tilde{\nu}<1800} f(\sigma\tilde{\nu})^2 d\tilde{\nu} \int_{\tilde{\nu}\geq 300}^{\tilde{\nu}<1800} g(\tilde{\nu})^2 d\tilde{\nu}}} \tag{S1}$$

To obtain the contributions of separate spectral regions to the full overlap integral, the integral in the numerator was split into separate regions as follows,

$$S_{fg} = \frac{\int_{\tilde{\nu}\geq 300}^{\tilde{\nu}<800} f(\sigma\tilde{\nu})g(\tilde{\nu})d\tilde{\nu} + \int_{\tilde{\nu}\geq 800}^{\tilde{\nu}<1200} f(\sigma\tilde{\nu})g(\tilde{\nu})d\tilde{\nu} + \cdots + \int_{\tilde{\nu}\geq 1400}^{\tilde{\nu}\leq 1800} f(\sigma\tilde{\nu})g(\tilde{\nu})d\tilde{\nu}}{\sqrt{\int_{\tilde{\nu}\geq 300}^{\tilde{\nu}\leq 1800} f(\sigma\tilde{\nu})^2 d\tilde{\nu} \int_{\tilde{\nu}\geq 300}^{\tilde{\nu}\leq 1800} g(\tilde{\nu})^2 d\tilde{\nu}}} \tag{S2}$$

with:

$\sigma = [\delta(\tilde{\nu} \leq 1580)\sigma_{\tilde{\nu}\leq 1580} + (1 - \delta(\tilde{\nu} \leq 1580)\sigma_{\tilde{\nu}>1580})\ ]$

$\sigma_{\tilde{\nu}\leq 1580} = 0.987$ and $0.970 \geq \sigma_{\tilde{\nu}>1580} \geq 0.900$

$\delta(\tilde{\nu} \leq 1580) = 1$ if $\tilde{\nu} \leq 1580$, else $\delta(\tilde{\nu} \leq 1580) = 0$

In our previous work,[1] we found that a global scaling factor σ of 0.987 gives a good comparison between experiment and theory. The amide I region requires, however, a different scaling factor. Both the basis set and hydration have an important influence on the calculated wavenumbers of the amide I vibrational modes. Therefore, the optimal scaling factor was first determined for the amide I region (Fig. S 6). By shifting the calculated amide I modes (1675-1800 cm^{-1}) with this scaling factor and the remainder of the spectrum by the global scaling factor of 0.987 prior to convoluting the line spectrum with the Lorentzian functions, a good alignment between experiment and theory is found (see Figure 11 and Figure 14). Optimization of the global scaling factor for the different calculated spectra gave only minor differences (1-2

%) in the similarity values as the optimal factor was found between 0.985-0.991 in this study. Therefore, the fixed scaling factor of 0.987 was used here.

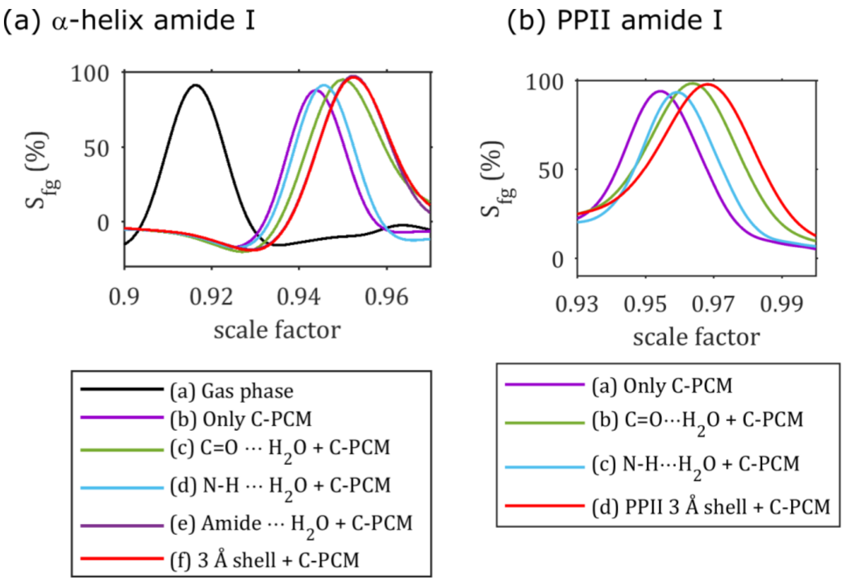


Fig. S 6 Determination of the optimal scaling factor of the amide I region of the calculated ROA spectra in Figure 11 (α-helix) and Figure 14 (PPII).

5 Standard deviations

Standard deviations σ were calculated using the following formula,

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2} \tag{S1}$$

with n the sample size, x_i the sample value of sample i and $\bar{x}$ the mean of the set of values.

6 Additional spectra

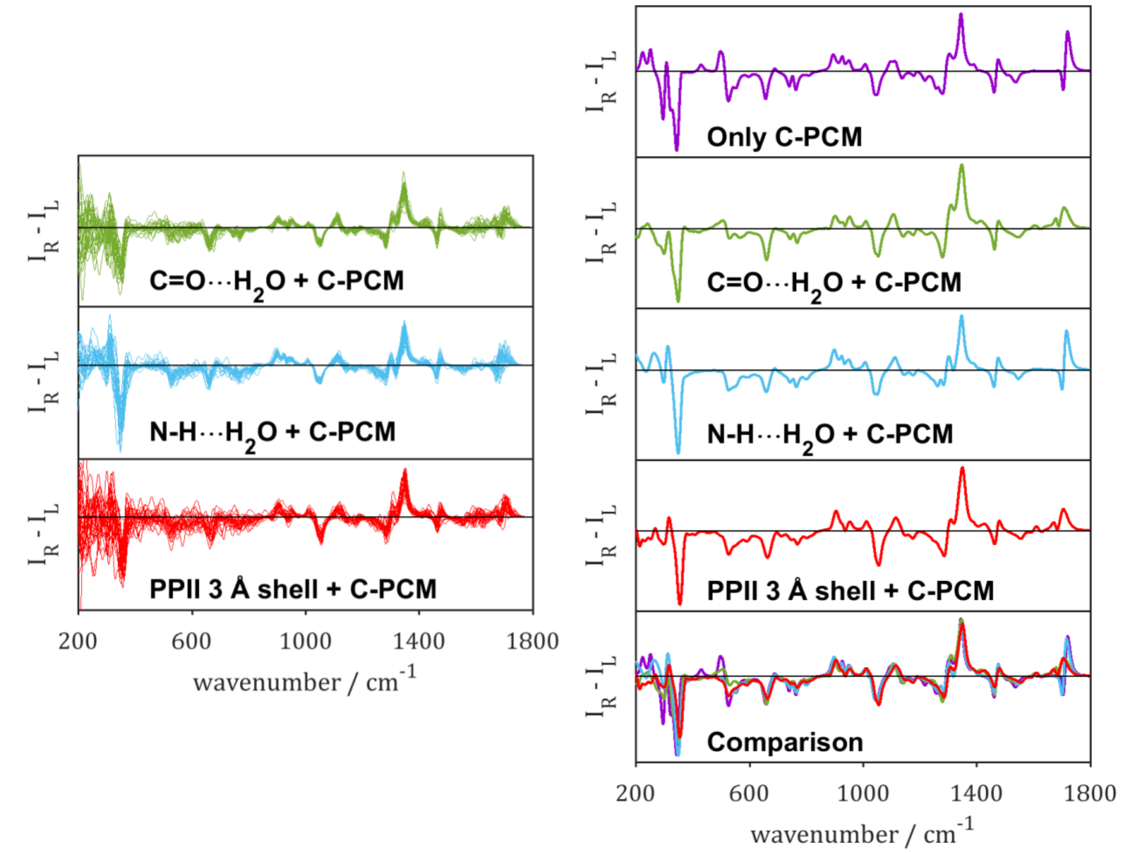


Fig. S 7 For the 30 hydration configurations of the α-helical conformation with $\phi;\psi \approx -59°;-44°$, the 30 individual spectra are shown on the left-hand side depending on the hydrogen bonding. On the right-hand side the mean of these ROA spectra are compared to that when only C-PCM is included.

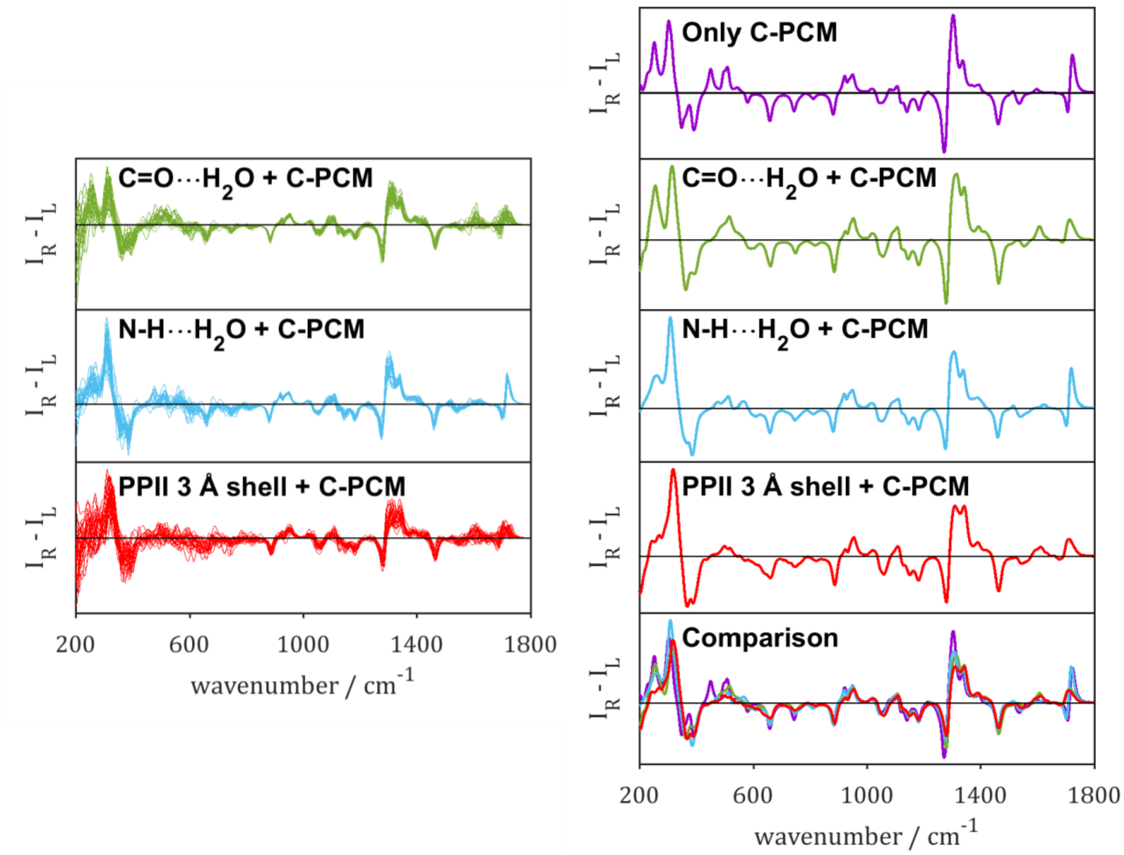


Fig. S 8 For the 30 hydration configurations of the α-helical conformation with ϕ;ψ ≈ -77°;-34°, the 30 individual spectra are shown on the left-hand side depending on the hydrogen bonding. On the right-hand side the mean of these ROA spectra are compared to that when only C-PCM is included.