Supplementary information

Ramachandran Mapping of Peptide Conformation Using a Large Database of Calculated Raman and Raman Optical Activity Spectra

Carl Mensch,[a,b] Laurence D. Barron[c] and Christian Johannessen*[a]

[a] Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, B-2020

Antwerp, Belgium. E-mail: christian.johannessen@uantwerpen.be.

[b] Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), 9000 Ghent, Belgium.

[c] School of Chemistry, University of Glasgow, Joseph Black Building, Glasgow G12 8QQ, United Kingdom.

Table of Contents **Page**

S1. Database construction: model selection and geometry optimization

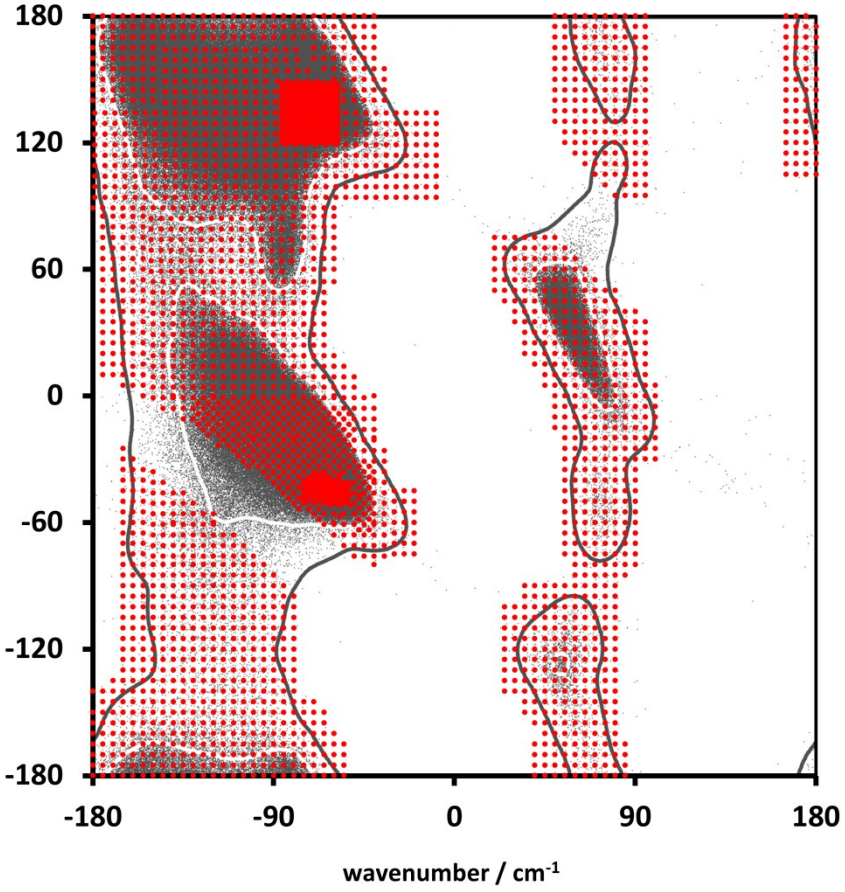


Figure S 1 Selection of 3035 HCO-$(Ala)_{11}$-$CONH_2$ model structures with repetitive φ and ψ angles (red dots) selected across the preferential and allowed regions of the Ramachandran plot. The region below the α-helical and π-helix region is for example not populated as the $(Ala)_{11}$ peptides cannot be created using repetitive angles; these geometries overlap internally.

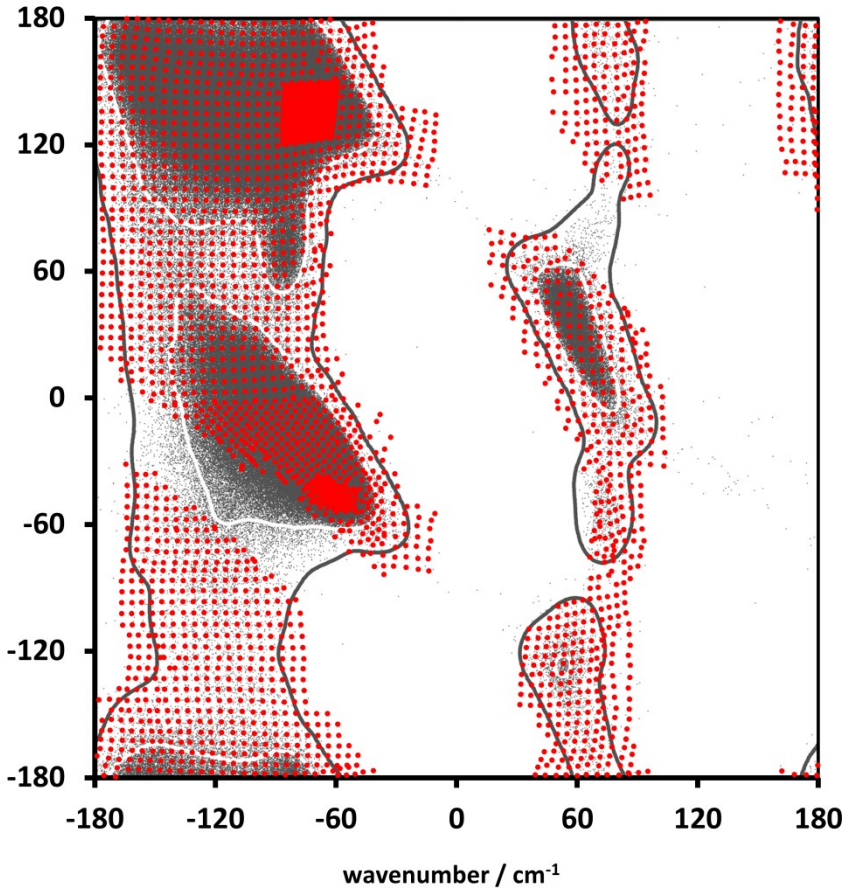


Figure S 2 Average φ and ψ angles for each of the 2902 optimized HCO-(Ala)$_{11}$-CONH$_2$ model structures in the final database.

S2. Calculation of Raman and ROA intensities

The Raman and Raman optical activity (ROA) intensities in the database are calculated as described in the main text by using Gaussian 09 rev. D.01. The spectrum S at each frequency ω can be calculated for each transition of frequency ω_i using the following formula:

$$S(\omega) = I\left[1 - \exp\left(-\frac{\omega_i}{k_B T}\right)\right]^{-1} \frac{1}{\omega_i}\left[4\left(\frac{\omega - \omega_i}{\Delta}\right)^2 + 1\right]^{-1}$$

This equation takes the temperature T (300 K) into account using a Boltzmann distribution factor with the Boltzmann constant k_B.[1] The spectral line shapes were represented as a Lorentzian function for each transition with a full width at half height Δ of 20 cm^{-1}. In this equation I is defined as:

$$I = 6\left(7\alpha_{ij}\alpha_{ij} + \alpha_{ii}\alpha_{ii}\right)$$

for Raman and the ROA backscattered (180°) intensities are calculated as:

$$I = \frac{48}{c}\left(3\alpha_{ij}G'_{ij} - \alpha_{ii}G'_{jj} + \frac{\omega_{exc}\varepsilon_{ijk}\alpha_{il}A_{jkl}}{3}\right)$$

where α, G' and A are the electric dipole-electric dipole, magnetic dipole-electric dipole, and electric quadrupole-electric dipole polarizability normal mode derivatives; ω_{exc} is the laser light frequency (532 nm); c is the speed of light and ε_{ijk} is an element of the antisymmetric Levi-Civita tensor.[2]

S3. Calculation of similarity index S_{fg}

To objectively assign experimental spectra using the spectral database, a robust procedure with as little user interference as possible is desirable. When using the S_{fg} similarity measure, two corrections are introduced: (1) A scaling factor needs to be introduced to correct for the global wavenumber shift and (2) a weighting function is used to take local shifts in the spectrum into account.

S3.1 Global wavenumber shift: Optimal scale factor

Because of *i.a.* the harmonic approximation, the simulated spectra are typically overestimated in the wavenumber dimension. To be able to reliably calculate the similarity index, the theoretical spectra first need to be shifted to be aligned with the experimental spectrum. To determine a general applicable scaling factor (SF) to shift the spectra in the wavenumber dimension, the optimal SF for three experimental spectra was determined. As the low wavenumber region has high intensity in the simulated spectra for single conformers and this region is unreliable because of the constrained geometry optimization in normal coordinates, the SF is calculated starting from 450 cm^{-1} as a minimum value (*vide infra)*.

Figure S 3 shows the arbitrary selection of five database spectra for evaluating the optimal scaling factors for the theoretical spectra compared to the experimental spectrum. These spectra were selected within the regions that contain the spectra that resemble the corresponding experimental spectrum the most.

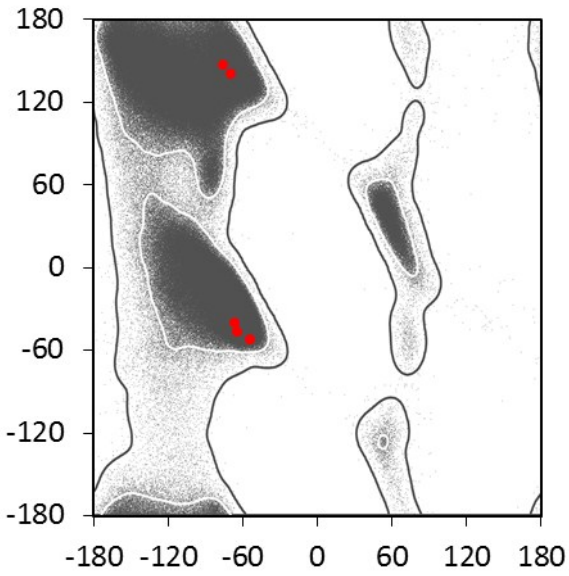


Figure S 3 Selection of 5 database spectra for evaluating the optimal scaling factors. Average Φ and Ψ angles are shown for each structure: -64.0°; -47.07° and -66.37°; -40.7° are assigned to the experimental AK21 spectrum (Figure S 4), -53.27°; -52.79° to PLGA at low pH (Figure S 8) and -69.06°; 139.72° and -57.83°; 146.73° to XAO (Figure S 7).

In Figure S 4 the spectra of two database spectra that reproduce the experimental spectrum of AK21 very well are shown, for example. Both spectra describe the -/+/+ patterns in the skeletal stretch and amide III regions quite good as well as the amide I -/+ couplet. Nonetheless, the amide I couplet is for example shifted to higher wavenumbers for the database spectrum with average dihedral angles Φ/Ψ of -64.34°; -47.07° compared to the spectrum of the -66.34°; -40.70° model. Therefore, the optimal scaling factor for the 450-1580 cm^{-1} and the 1580-1700 cm^{-1} regions are given separately in Figure S 5 and Figure S 6, respectively.

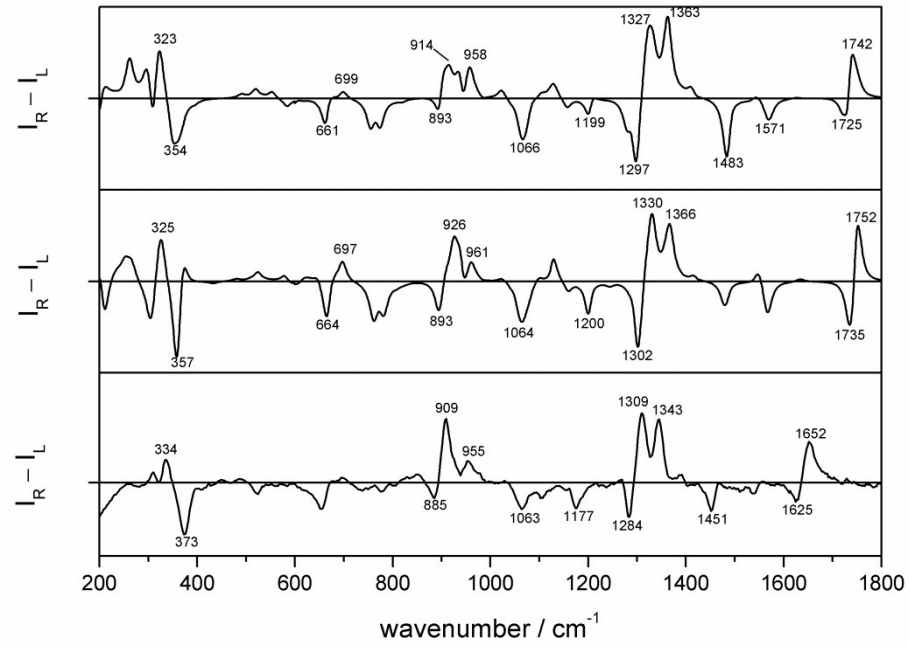


Figure S 4 Two unscaled calculated ROA spectra of the database (top -66.34°; -40.70° and middle -64.34°; -47.07°) compared to the experimental spectrum of AK21 (bottom).

In Figure S 5, the dependence of the similarity index S_{fg} on the scale factor is depicted for the five database spectra compared to the experimental spectrum that can be assigned to them. For each theoretical spectrum, the S_{fg} curve compared to the scaling factor shows the expected behaviour with a maximum value for acceptable scaling factors. For the AK21 example, we see that the optimal scaling factor for the 450-1580 cm^{-1} range is not exactly the same for both theoretical spectra, but it is similar.

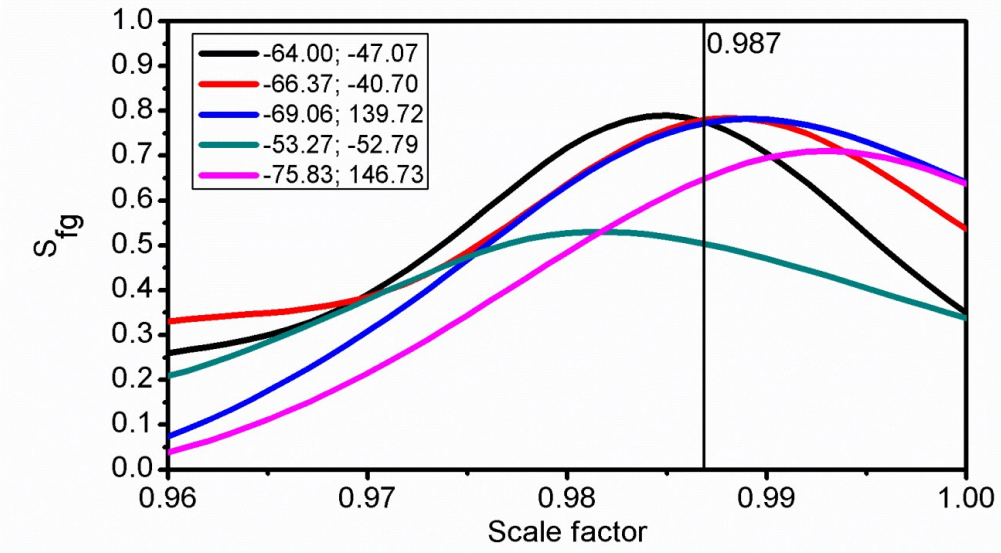


Figure S 5 Similarity index S_{fg} for the wavenumber range 450-1580 cm^{-1} as a function of the scaling factor for five different conformations in the database. The database structures -64.00°; -47.07° and -66.37°; -40.70° are compared to the experimental AK21 spectrum, -53.27°; -52.79° to PLGA (helix) and -69°.06°; 139.72° and -57.83°; 146.73° to XAO.

For the 1580-1700 cm^{-1} region, the S_{fg} is more crucially sensitive to the scaling factor, as can be seen from the narrow shaped curves in Figure S 6 which do not coincide. This is because only the amide I couplet accounts for the similarity in this region. In the two theoretical spectra given for AK21 in Figure S 4, the slight shift in the amide I position, results in the two optimal scaling factor of 0.940 and 0.946. If one of these factors is used to shift the amide I region of the other spectrum, the similarity would be significantly lowered as can be seen from Figure S 6. Or even worse, if the optimal scaling factors of the theoretical spectra that are assigned to the experimental spectrum of the XAO peptide (0.955 or 0.956) would be used to scale the amide I of the AK21 theoretical spectra, the similarity would be negative for the -64.00°; -47.07° spectrum and less than 0.2 for the -64.34°; -47.07° spectrum.

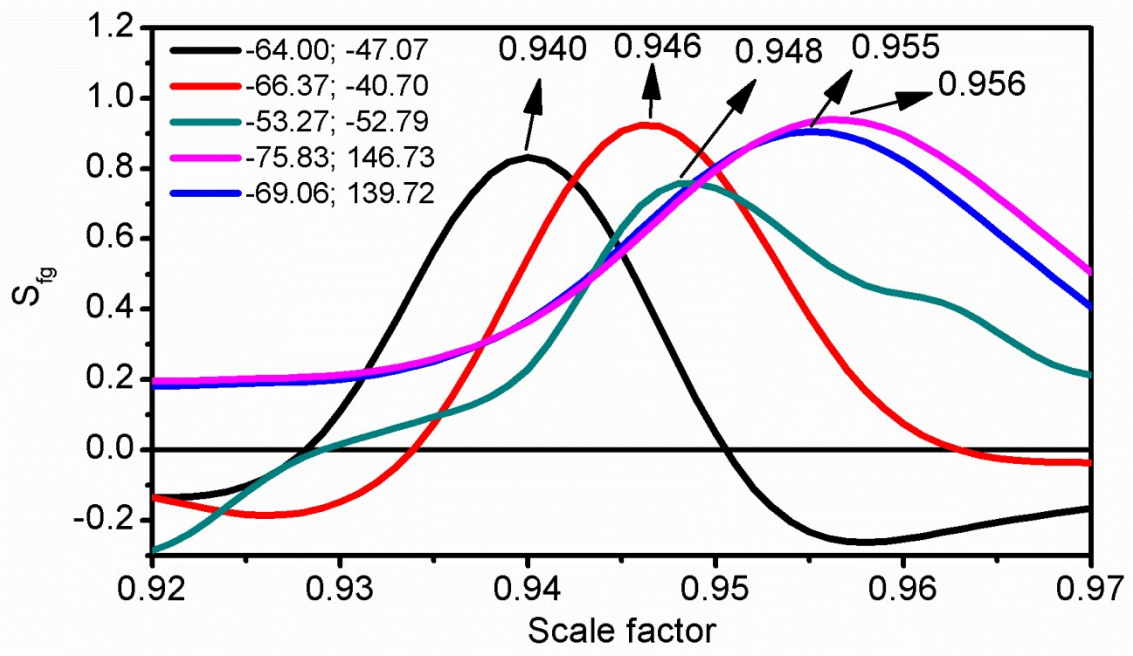


Figure S 6 Similarity index S_{fg} for the wavenumber range 1580-1750 cm^{-1} (amide I region) as a function of the scaling factor for five different conformations in the database.

As can be seen from Figure S 6 and Figure S 7, the amide I region of the database spectra assigned to the experimental ROA spectrum of the XAO peptide, have a higher optimal scaling factor compared to the amide I region of the AK21 spectra.

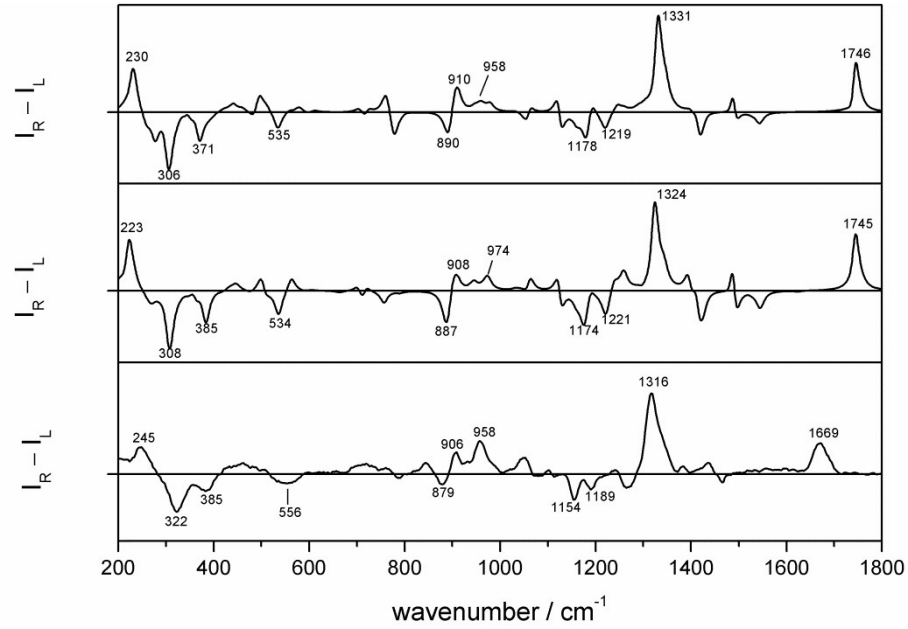


Figure S 7 Unscaled database spectra (top -69.06°; 139.72° and middle -57.83°; 146.73°) compared to the XAO experimental ROA spectrum (bottom).

As an example, the experimental spectrum of PLGA (at low pH) is compared to one theoretical spectrum (-53.27°; -52.79°). While the wavenumber scaling will improve the overlap of the amide III region, small shifts in the skeletal stretch and amide III region occur and result in a lower S_{fg} than a visual comparison would suggest. To account for this, a weighting function can be applied to the data before calculating the S_{fg} (S3.2).

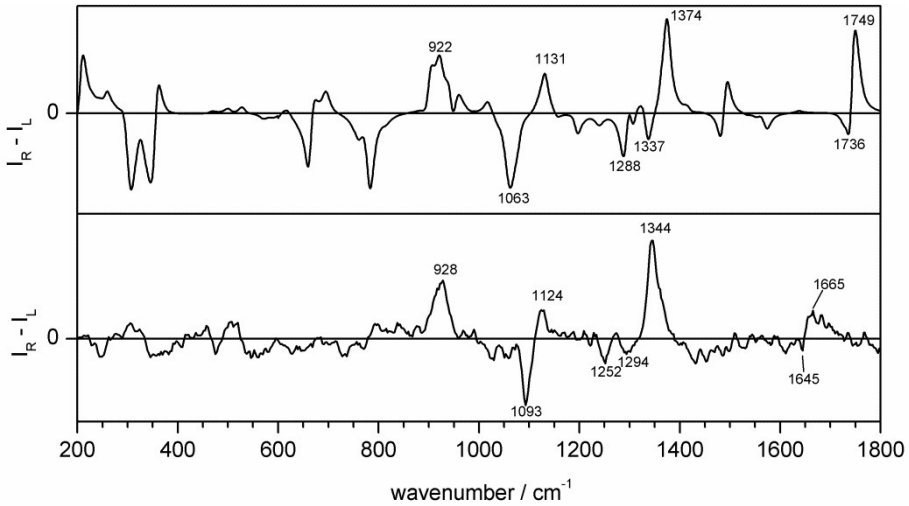


Figure S 8 Unscaled database spectrum of the database model with average Φ/Ψ angles of -53.27°; -52.79° (top) compared to the PLGA (helix) experimental ROA spectrum.

To conclude, a global wavenumber scaling factor of 0.987 seems to be valid for the calculated spectra in the database. However, the amide I region requires a lower scaling factor to account for the wavenumber shift. Since this region is a key spectral region for peptides and proteins, it is included here using a separate scaling factor. Yet, as can be seen from Figure S 6, the scaling factor needs to be optimized to properly align this region with the same region in the experimental spectrum. To calculate the similarity index for one data point in the database, firstly the optimal scaling factor SF_I is determined only for the amide I region 1580-1750 cm^{-1} in steps of 0.001 for the SF_I ranging from 0.930-0.970. Next, these amide I normal modes are shifted separately from the other normal modes using the scale factor $SF_I/0.987$. Finally, the entire calculated spectrum is scaled with the factor 0.987 and the similarity is calculated for the entire range.

S3.2 Local shifts: Triangular weighting

The SF of 0.987 is chosen to correct for a general shift in the wavenumber dimension. However, small local shifts in the spectrum can occur, resulting in a reduced S_{fg} (see e.g. the couplet in the skeletal stretch regions of PBLG and PLGA). Compared to a visual assignment of the spectrum, the similarity measure might then present lower values than expected. In similarity measures that are used in chiroptical spectroscopy, a triangular weighting function is sometimes used to correct the similarity value for local shifts by taking the neighbouring spectral points into account for each point.[3] A spectrum g is weighted according to the following formulas using the triangular weighting function w. In this work a weighting function of 40 cm^{-1} is used.

$$g(v_0) \leftarrow \int g(v)w(v - v_0)dv$$

$$w(v - v_0) = 1 - \frac{|v - v_0|}{l} for\ |v - v_0| \leq l$$

$$w(v - v_0) = 0 \; for \; |v - v_0| > l$$

This weighting is applied both to the calculated spectrum *f* and the experimental spectrum *g*. The result of the weighting on the similarity mapping is shown in Figure S 9, which shows the comparison of the similarity mapping with and without the triangular weighting of 40 cm^{-1} included.

S3.3 Final S_{fg} formula

Finally, the calculation of the similarity index S_{fg} that compares the (weighted) *f* with *g* for the wavenumber range 450-1800 cm^{-1} can be summarised as:

$$S_{fg} = 100\% \cdot \frac{\int_{450}^{1800} f(\sigma\tilde{\nu}) g(\tilde{\nu}) d\tilde{\nu}}{\sqrt{\int_{450}^{1800} f(\sigma\tilde{\nu})^2 d\tilde{\nu} \int_{450}^{1800} g(\tilde{\nu})^2 d\tilde{\nu}}}$$

with:

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

$$\sigma_{\tilde{\nu} \leq 1580} = 0.987 \; and \; 0.970 \geq \sigma_{\tilde{\nu} > 1580} \geq 0.930$$

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

S4. Similarity maps of AK21, XAO, PBLA, PLA, PBLG, PLGA and supplementary spectra

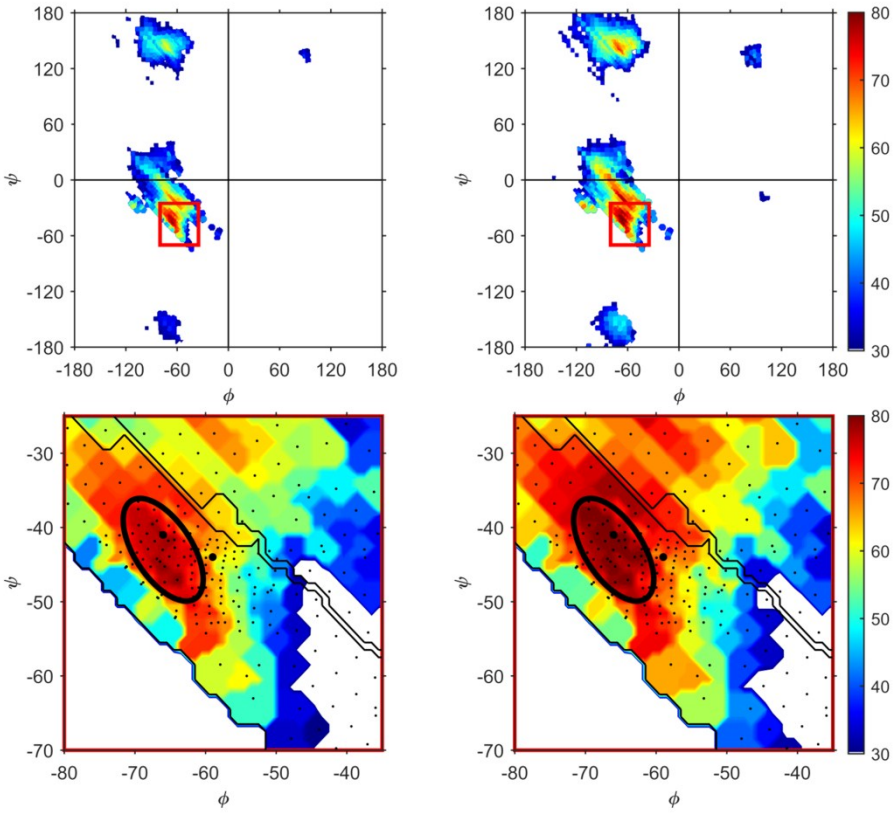
This section contains the similarity maps of all experimental spectra and shows the effect of the triangular weighting (window size of 40 cm^{-1}) in Figure S 8, followed by spectra that supplement the main paper. The black ellipses in the contour maps delineate the database Raman or ROA spectra f_i that are averaged and compared to the experimental counterparts. The average spectrum f_{ave} at each wavenumber $\tilde{\nu}$ is calculated using the following equation for the number of spectra *n* within the ellipse.

$$f_{ave}(\tilde{\nu}) = \frac{1}{n} \sum_{i=1}^{n} f_i(\tilde{\nu})$$

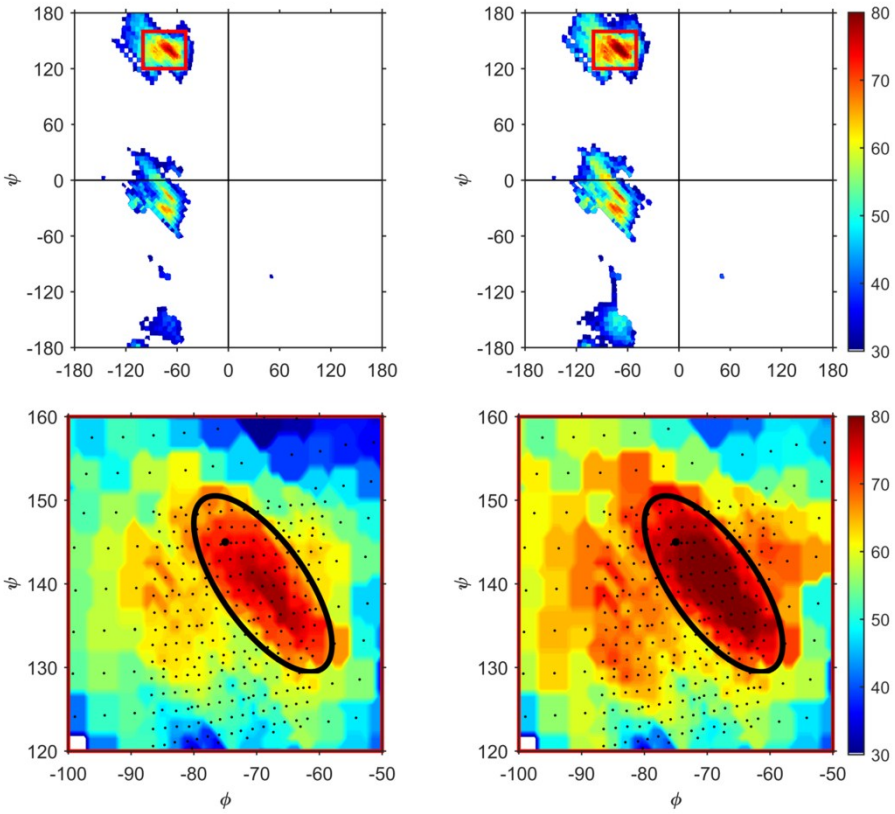
S4.1 Effect of triangular weighting on the similarity maps

The similarity maps both with and without triangular weighting (40 cm^{-1}) for all peptides in this study are depicted in Figure S 9.

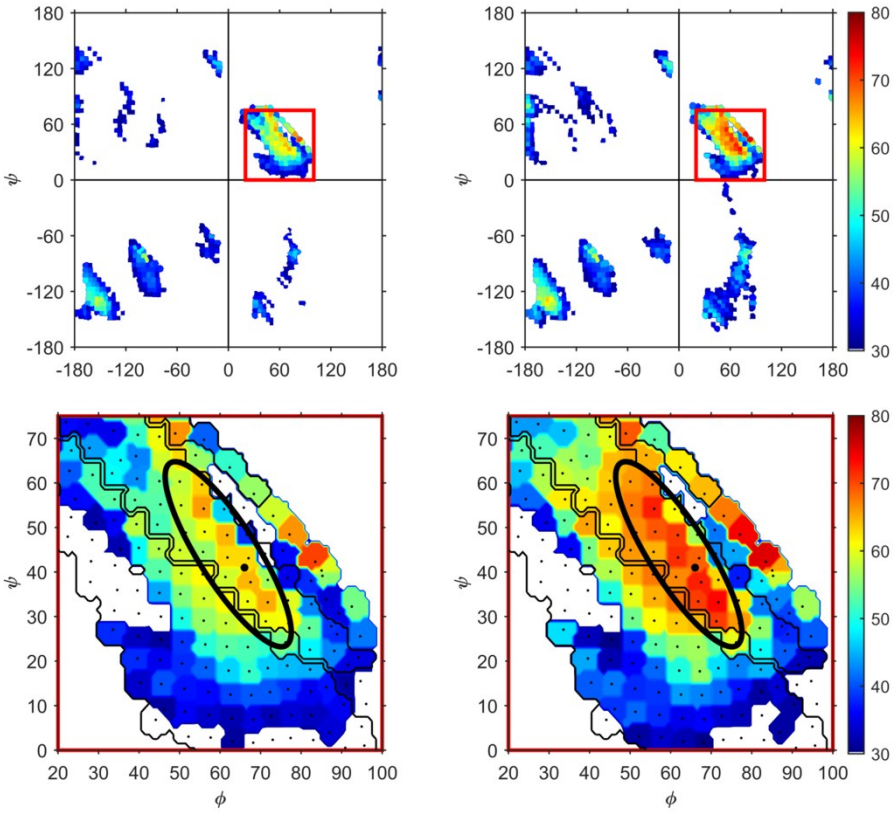
(1) AK21


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
ϕ
80
70
60
50
40
30
-30
-40
-50
-60
-70
-80 -70 -60 -50 -40

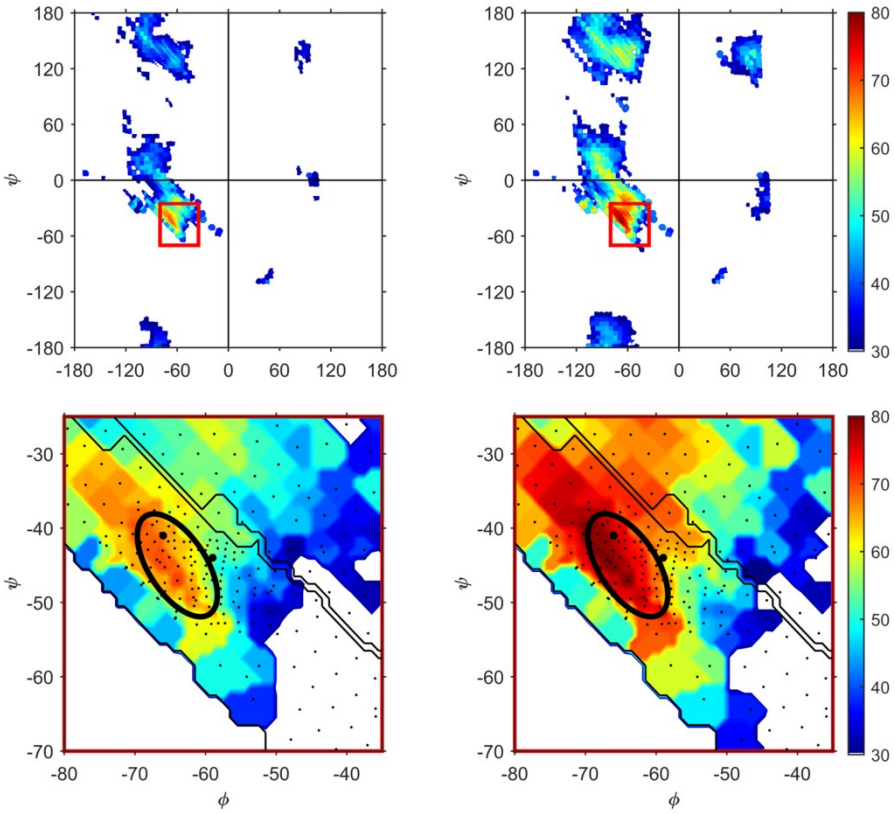
(2) XAO


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
ϕ
80
70
60
50
40
30
160
150
140
130
120
-100 -90 -80 -70 -60 -50

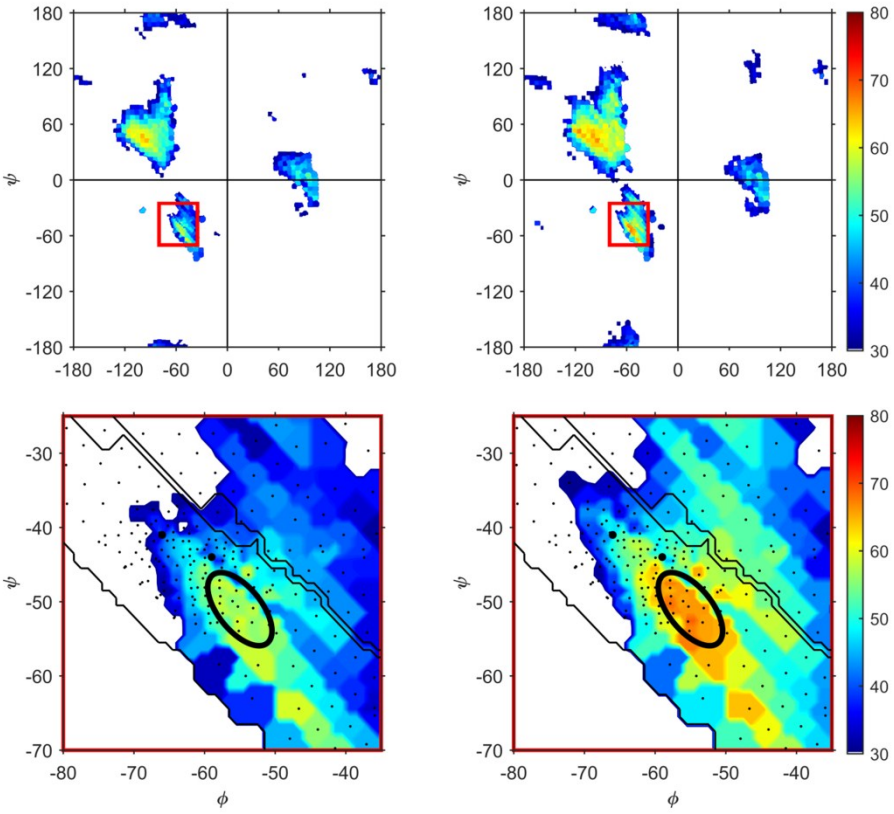
(3) PBLA


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
80
70
60
50
40
30
70
60
50
40
30
20
10
0
ϕ
20 30 40 50 60 70 80 90 100

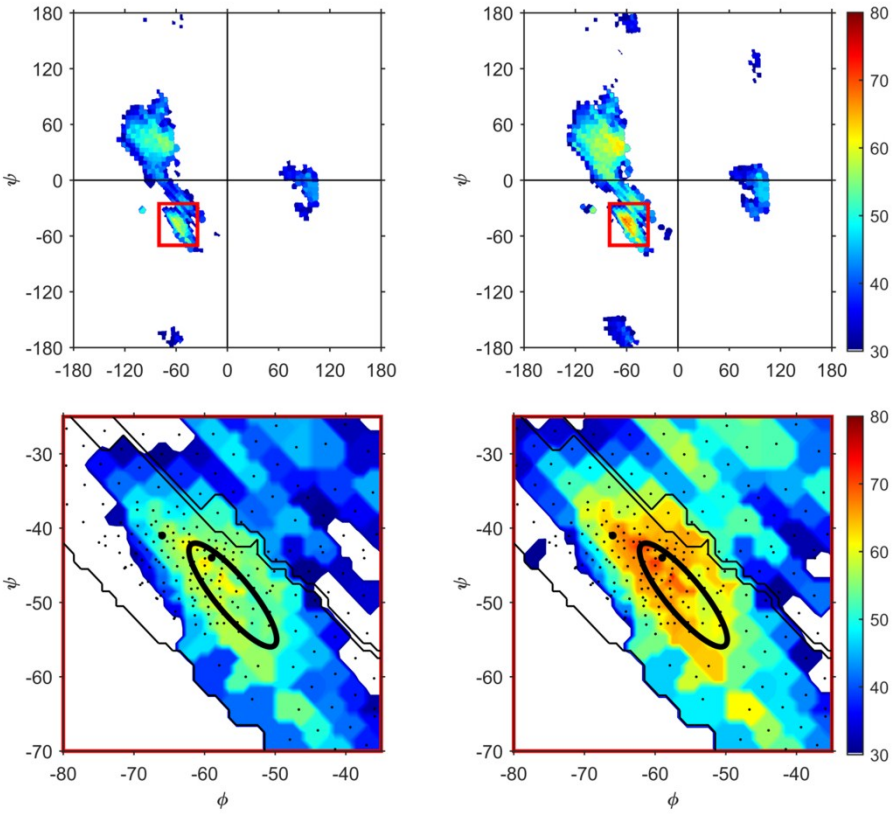
(4) PLA


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
80
70
60
50
40
30
-30
-40
-50
-60
-70
ϕ
-80 -70 -60 -50 -40

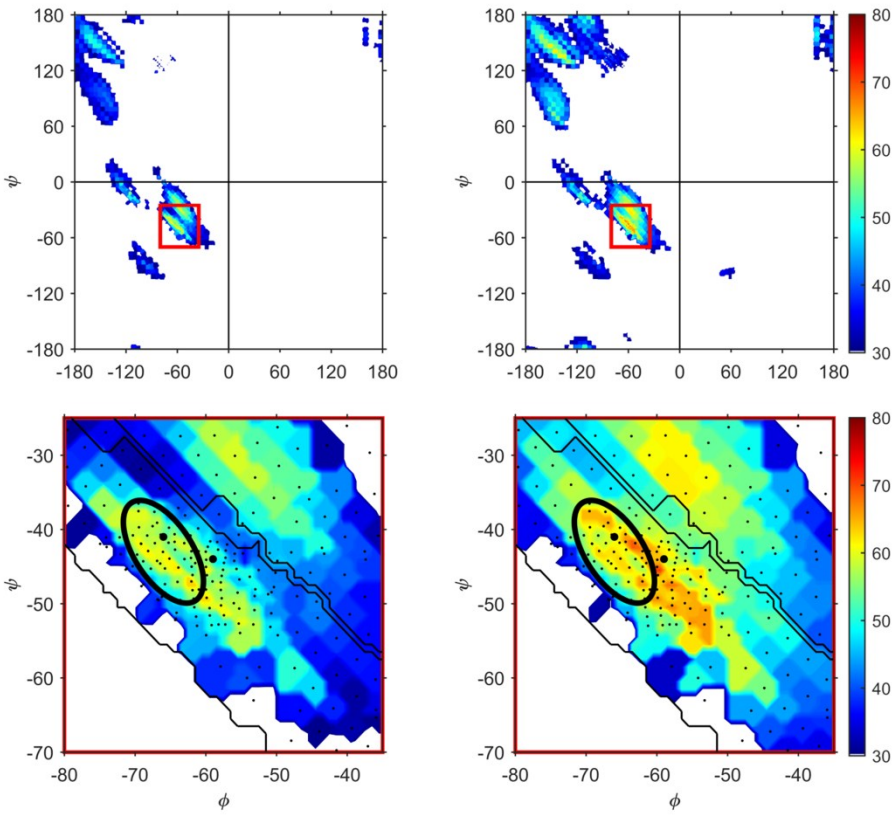
(5) PBLG


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
80
70
60
50
40
30
-30
-40
-50
-60
-70
-80 -70 -60 -50 -40
ϕ

(6) PLGA


180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
80
70
60
50
40
30
-30
-40
-50
-60
-70
-80 -70 -60 -50 -40
ϕ

(7) AK21 after H/D exchange



(8) XAO after H/D exchange

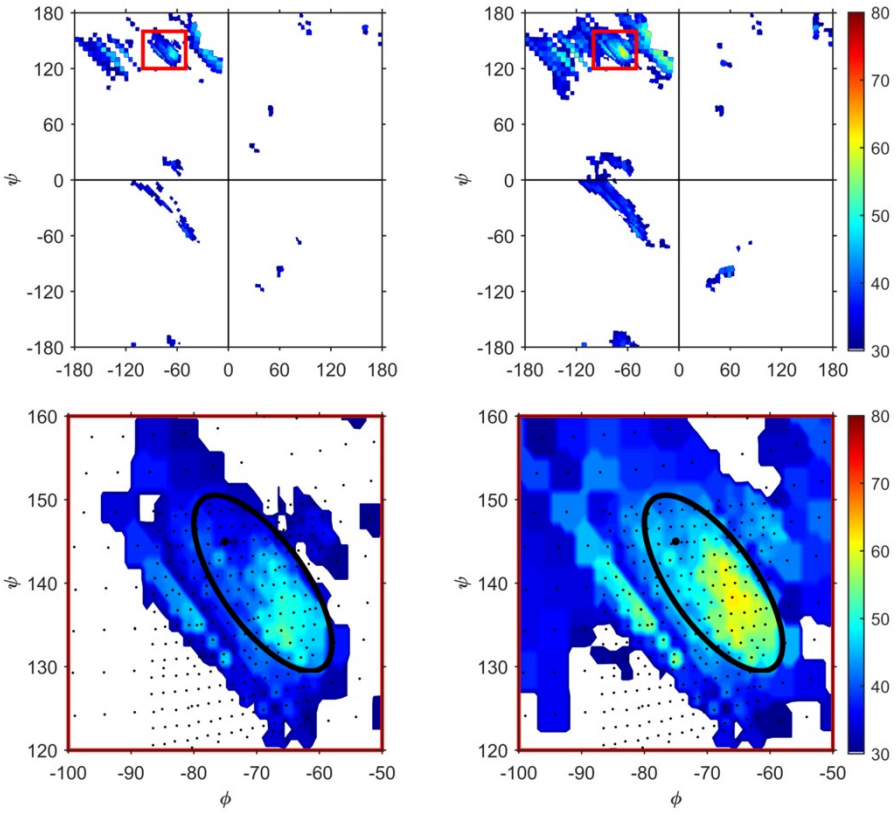


Figure S 9 Similarity maps of (1) AK21 (2) XAO (3) PBLA (4) PLA (5) PBLG (6) PLGA (7) AK21 after H/D exchange (8) XAO after H/D exchange without triangular weighting (left) and with triangular weighting of both the experimental and the database spectra using a triangular weighting function of 40 cm^{-1} (right). To take the conformational flexibility of the peptides in aqueous solution into account, the spectra within the manually selected ellipses are averaged and compared to the experimental spectrum and the database spectrum with the highest similarity (see figures 2 and 3). The similarity maps 1-6 use the original database; 7-8 are generated using the database where all the spectra are calculated using the H/D exchanged model structures. The similarities are calculated for the wavenumber range 450-1800 cm^{-1} unless when measured in $CHCl_3$ or DCA, the 800-1800 cm^{-1} region was used.

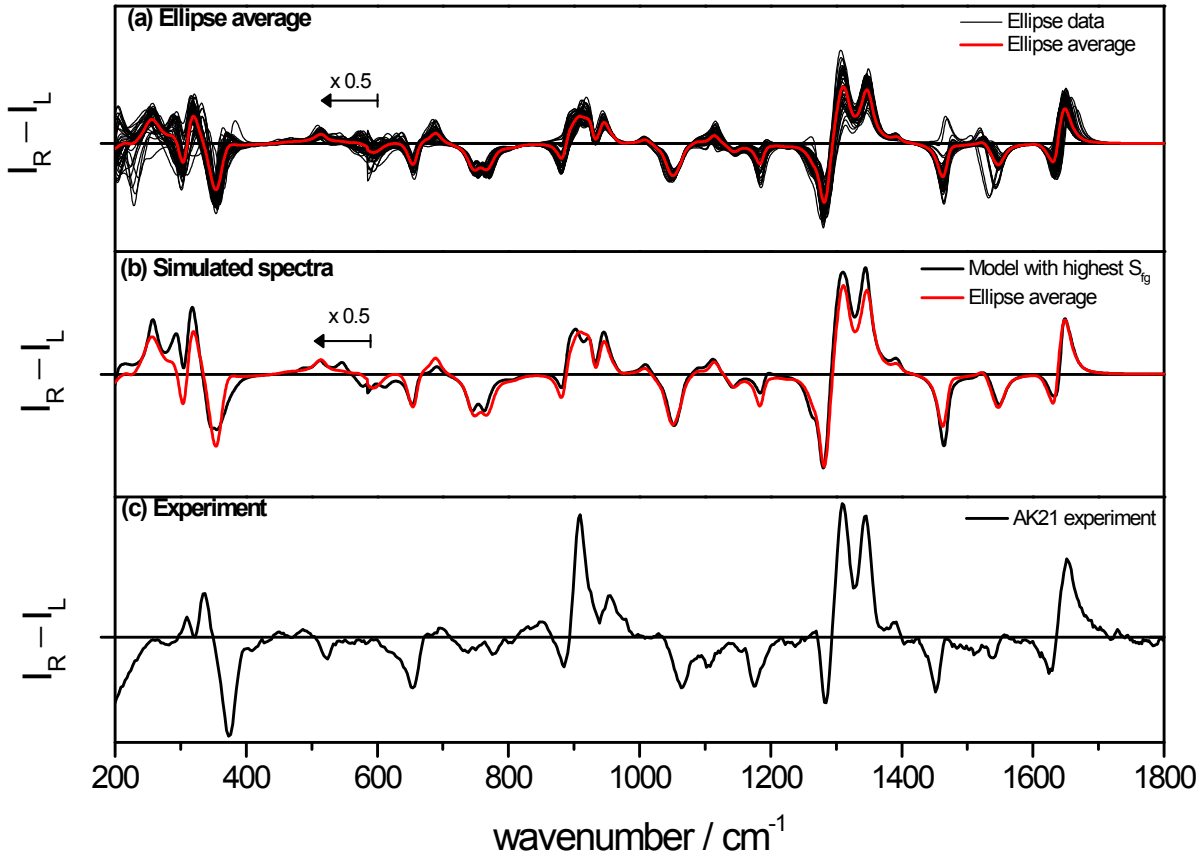


Figure S 10 ROA spectra of AK21: (a) spectra delineated by the ellipse in Figure S 9, (b, red) average spectrum of the ellipse spectra (b, black) spectrum with the highest S_{fg} compared to (c) the experimental spectrum.

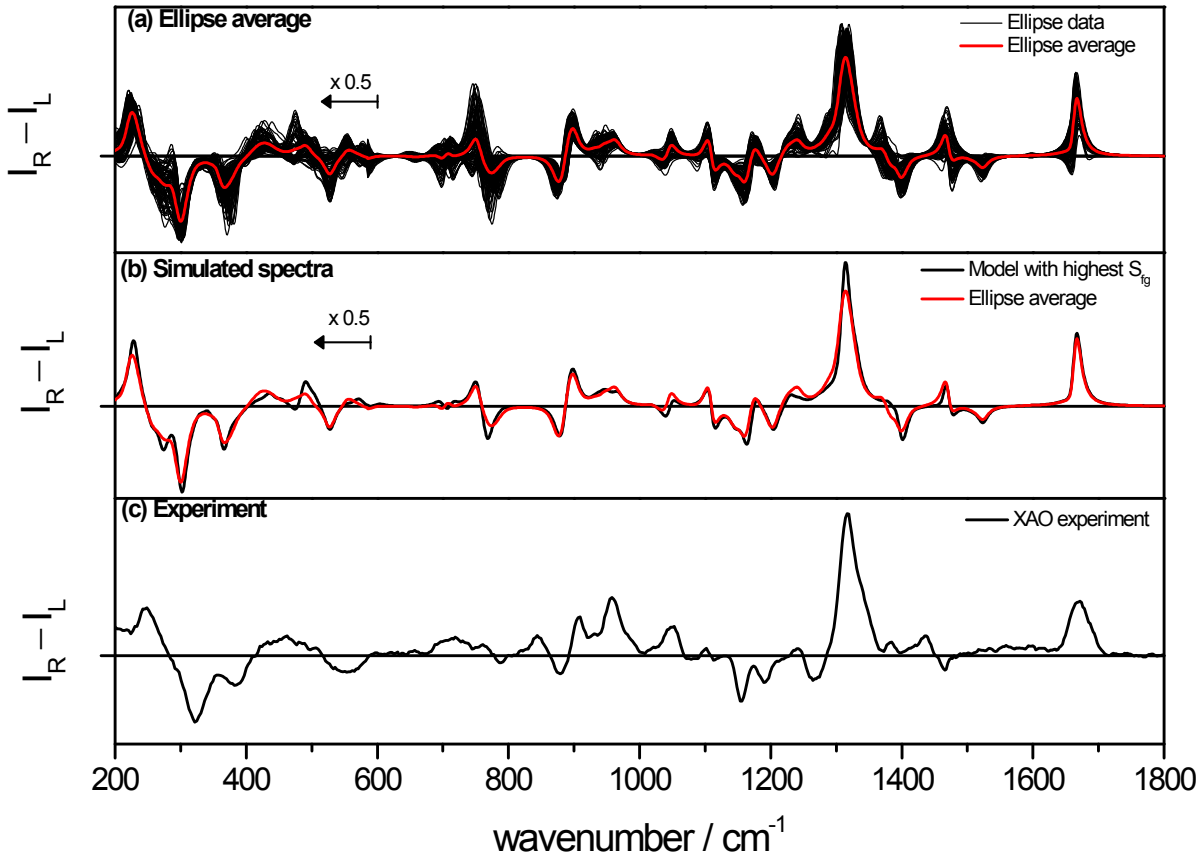


Figure S 11 ROA spectra of XAO: (a) spectra delineated by the ellipse in Figure S 9, (b, red) average spectrum of the ellipse spectra (b, black) spectrum with the highest S_{fg}. compared to (c) the experimental spectrum.

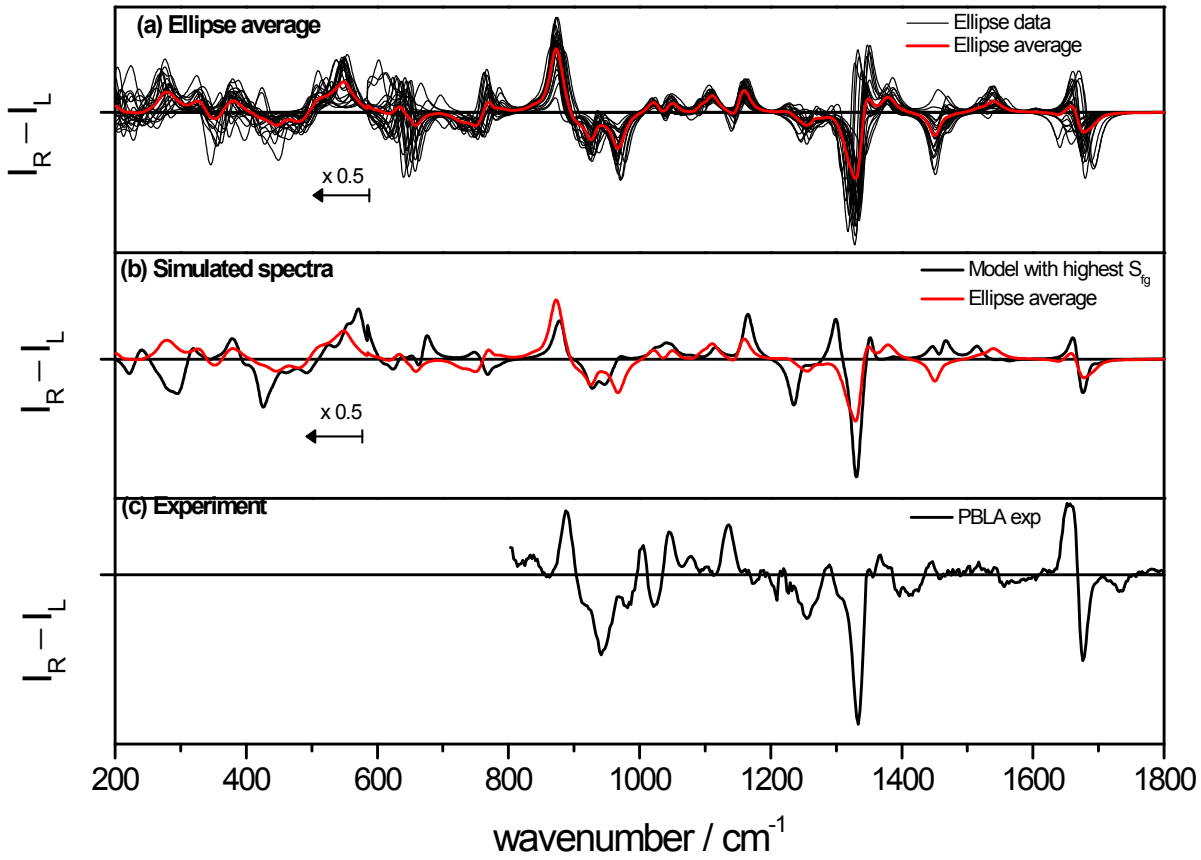


Figure S 12 ROA spectra of PBLA: (a) spectra delineated by the ellipse in Figure S 9, (b, red) average spectrum of the ellipse spectra (b, black) spectrum with the highest S_{fg}. compared to (c) the experimental spectrum.

S4.2 Using the wavenumber range 300–1800 cm^{-1}

In Figure S 9, the similarities are calculated using the 450–1800 cm^{-1}. As can be seen from Figure S 13, the similarity values would otherwise be dominated by the region below 450 cm^{-1}, because of the high intensities. Figure S 13 shows the similarity map of AK21, using the wavenumber range 300-1800 cm^{-1} and displays the highest similarity in the 3_{10} region. of the Ramachandran plot, while AK21 is assigned to α-helical structure.

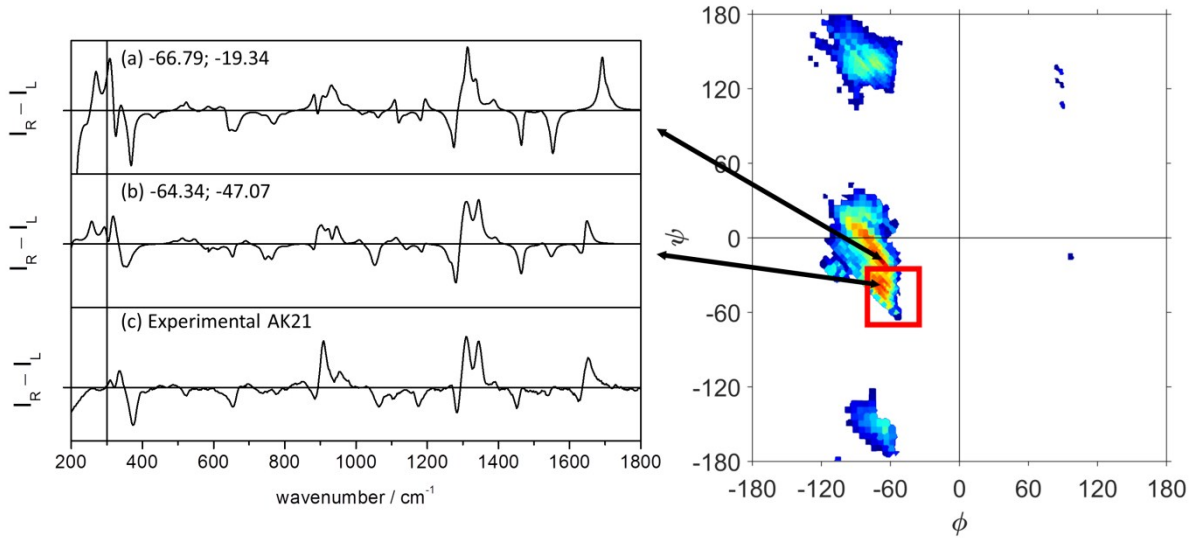


Figure S 13 The similarity map of AK21 generated using the wavenumber range 300–1800 cm^{-1} and a triangular weighting function of 40 cm^{-1}. shows the highest similarity for the database spectrum (a) for a model with average Φ;Ψ angles of -66.79°; -19.37°, while the spectra in the α-helical spectrum display lower similarities, for example for the spectrum of the database model with average angles -63.34°; -47.07°. (c) the experimental ROA spectrum of AK21.Spectrum (a) and (b) are plotted using the same y-axis setting.

S5. Coincidental spectral overlap in similarity maps

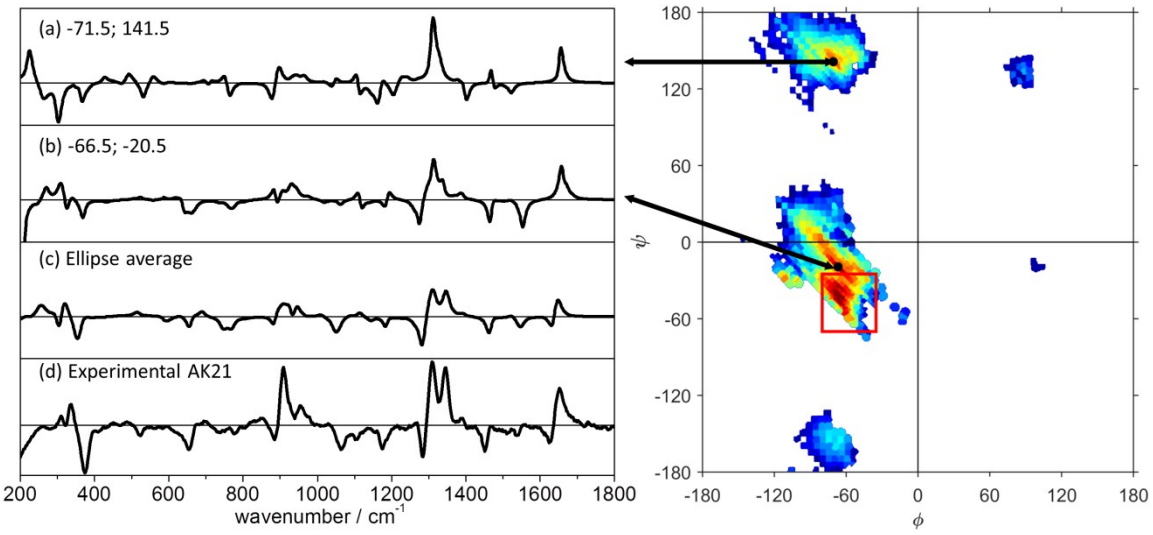


Figure S 14 ROA spectra of (a) the database model with average Φ/Ψ angles -71.5°/141.5° and (b) -66.5°/-20.5°, (c) the ellipse average as shown in figure 2 and (d) the experimental spectrum of AK21.

S6. Hydrogen bonding pattern indication in similarity maps

By evaluating the hydrogen bonding properties of the structures included in the database, the secondary structure types can easily be visualised as different regions of the Ramachandran plot.

The implemented hydrogen bond assessment is based on five geometric criteria[4]:

(1) The maximum distance between the amide oxygen and hydrogen is set at 2.5 Å.
(2) The maximum distance between the amide oxygen and nitrogen is set at 3.9 Å.
(3) The angle formed by oxygen, hydrogen and nitrogen "OHN" is limited to 90°<OHN<180°.
(4) The angle formed by the α-carbon, oxygen and nitrogen "CON" is limited to 90°<CON<180°.
(5) The angle formed by the α-carbon, oxygen and hydrogen "COH" is limited to 90°<COH<180°.

In Figure S 15 the three angular criteria are depicted.

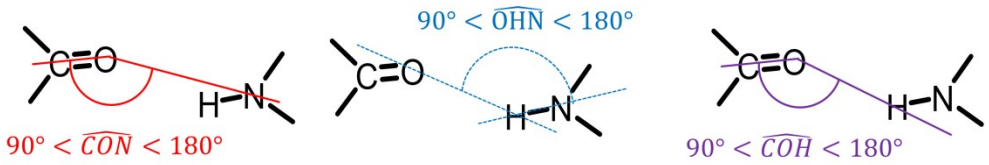


Figure S 15 The three angular criteria used in the hydrogen bond assessment

If the hydrogen bonds in a peptide model in the database fit all five requirements, the conformational type is specified as indicated in Table S 1. The hydrogen bonding type is indicated in the Ramachandran plots in the main section of the paper as a black solid contour line that groups the structures with the same hydrogen bonding type. The different regions in the database are indicated in different colours in Figure S 16.

Table S 1 Hydrogen bonding in the peptide database

Intrachain hydrogen bonding	Secondary structure	Colour in Figure S 16
i+2 → i	2_7 helix γ-helix/turn	Blue
i+3 → i	3_{10} helix	Green
i+4 → i	3.6_{13} helix / α-helix	Yellow
i+5 → i	4.1_{16} helix / π-helix	Red
i → i + 6 (inverse direction)	/	Brown

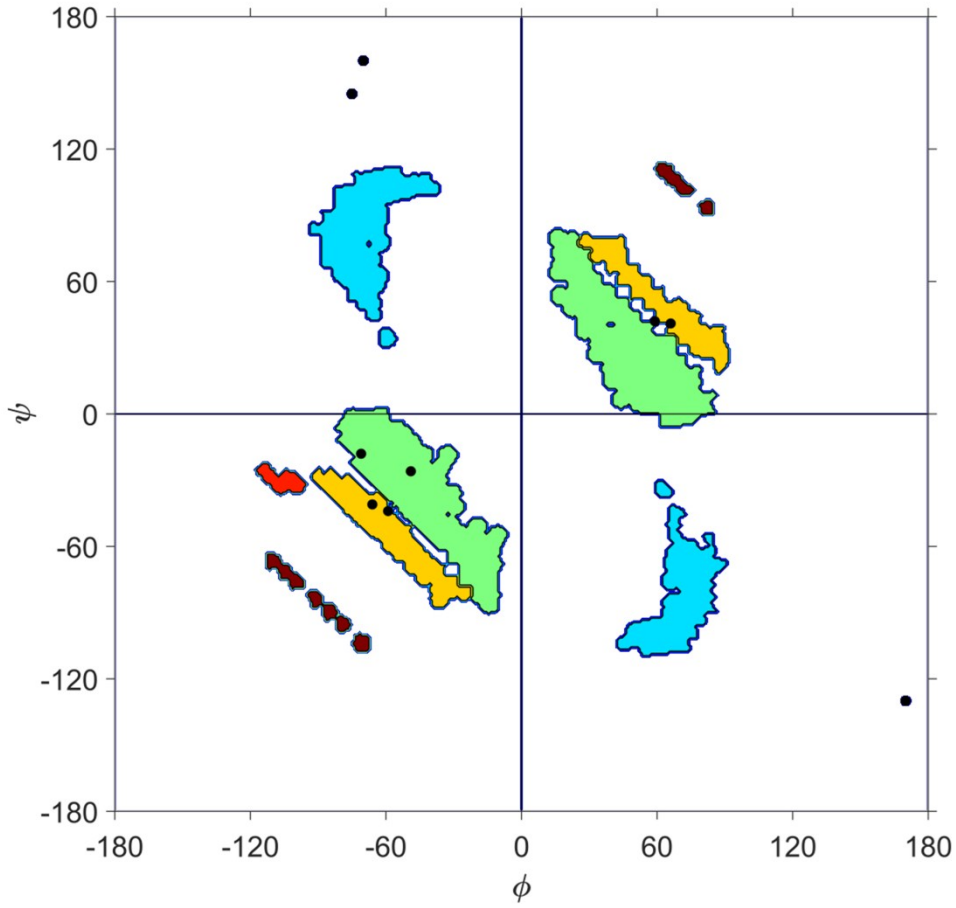


Figure S 16 Different structural regions in the Ramachandran plot based on hydrogen bonding. See Table S 1 for the colour coding. Both the right- and left-handed alternatives are indicated in the same colour (i+n → i and i → i+n).

S7. Effect of hydrogen bond orientation on the amide III region

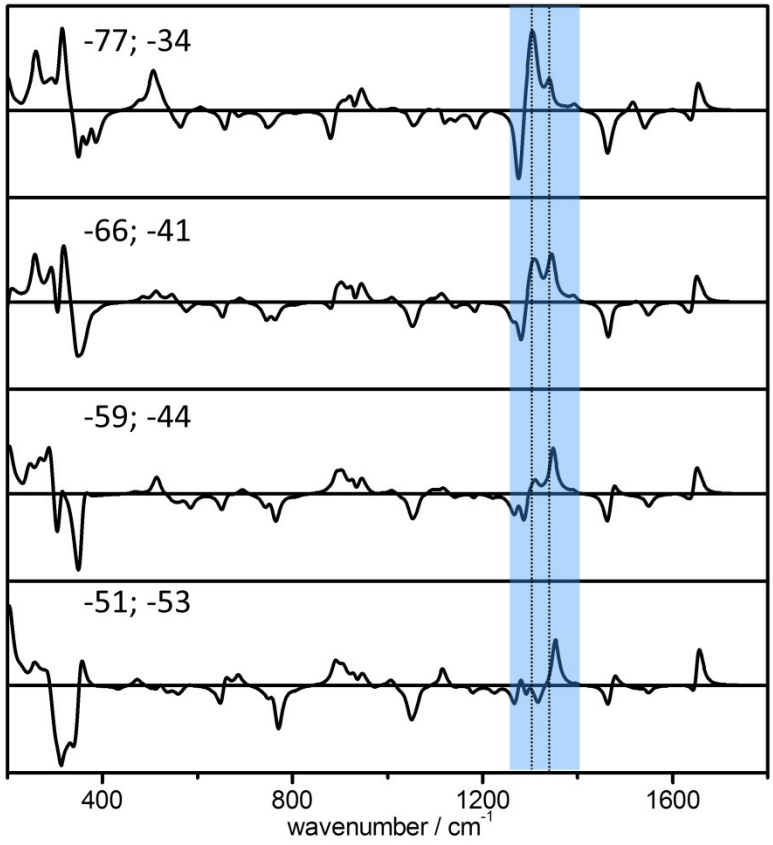


Figure S 17 Four different database ROA spectra in the α-helical region that demonstrate the effect of different conformations and thus hydrogen bond orientations on the amide III spectral region. See Figure 4 in the main text.

S8. Effect of partial deuteration: 1300/1340 cm^{-1} ratio

To demonstrate the effect of partial deuteration on the 1300/1340 cm^{-1} ratio of the two α-helical bands in the amide III region of the spectrum, the 13 protons in two model structures in the database were firstly one by one exchanged for deuterons starting from the C-terminal to the N-terminal after which the ROA spectra were calculated (Figure S 17).

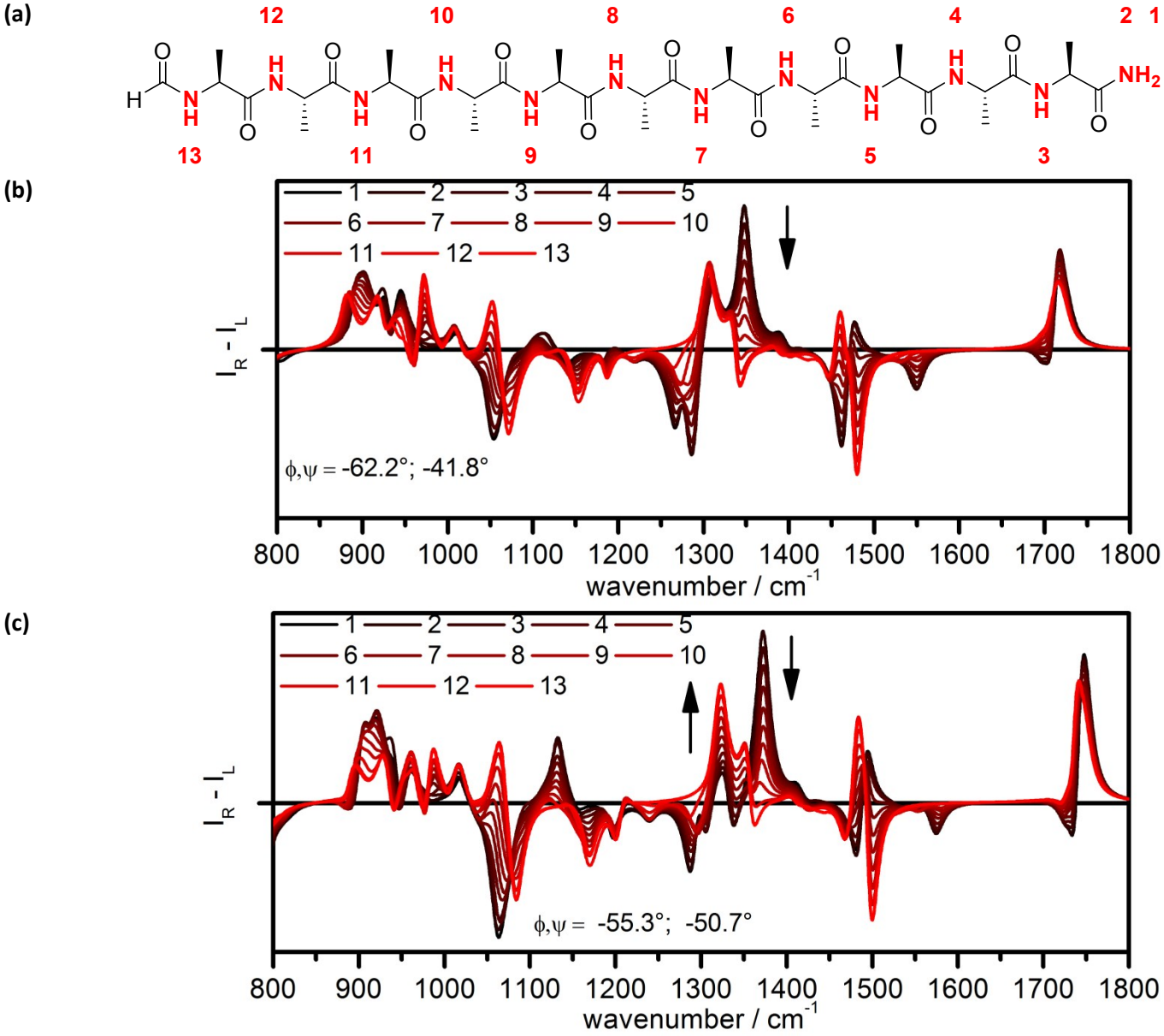


Figure S 18 ROA spectra of two database structures after progressive deuteration: From black to red, (a) the N-H protons are one after the other exchanged by deuterons starting from the C-terminus to the N-terminus until all 13 N-H's are exchanged. Model structures with average dihedral angles (b) -62.2°; -41.8° and (c) -55.3°; -50.7°.

Next, for the same two database structures, the four N-H protons as depicted in Figure S 18 were replaced by deuterons to simulate the effect of partial deuteration on one side of an α-helix. The ROA spectra are depicted in Figure S 18.

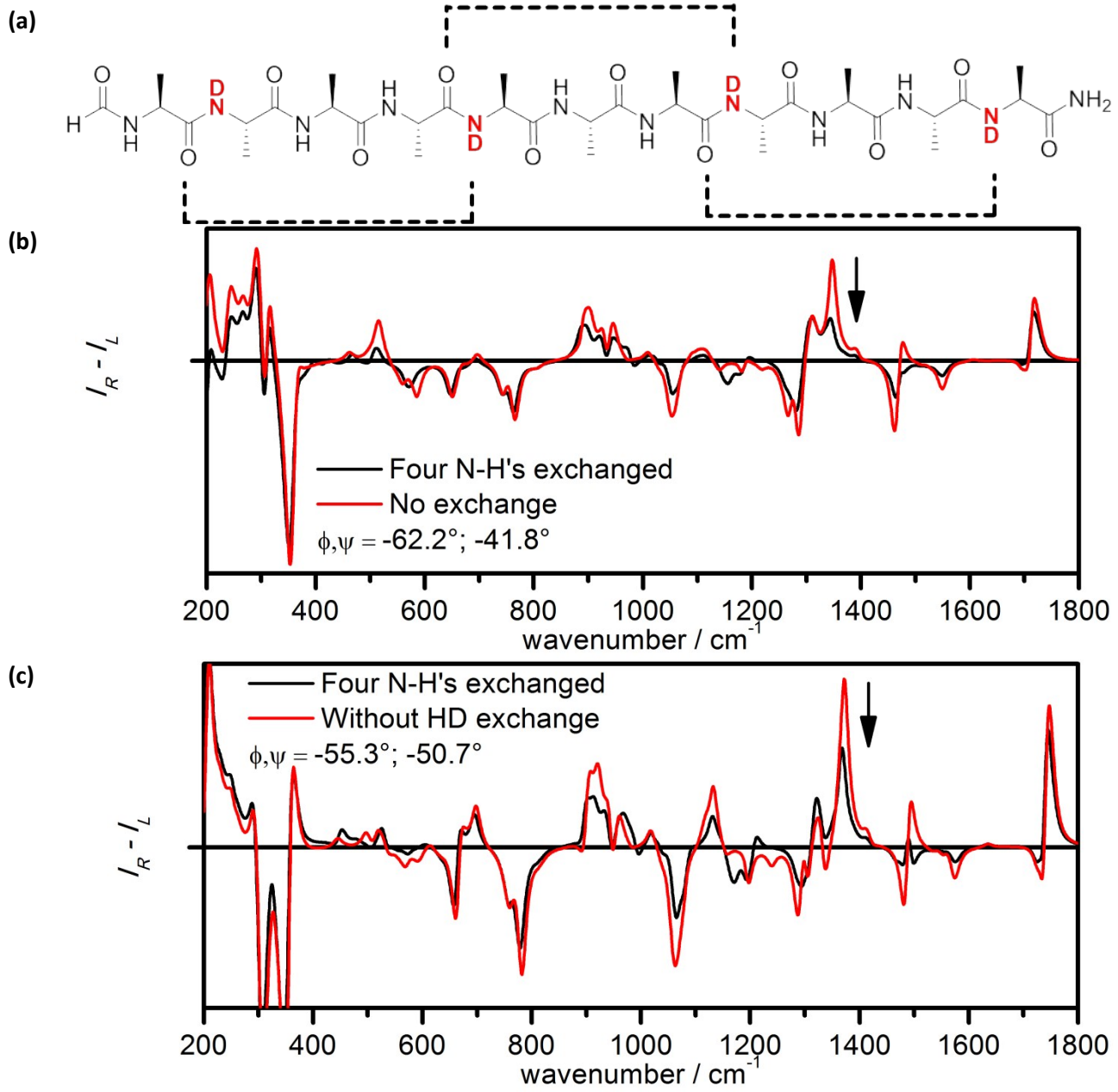


Figure S 19 (a) Exchange of four N-H protons by deuterons in the backbone of two database structures with average backbone angles (b) -62.2; -41.8° and (c) -55.3°; -50.7°, the ROA spectra were calculated (black) and compared to the non-deuterated ones (red).

S9. Ramachandran plot of insulin

The Ramachandran plot of insulin of human and bovine insulin for both $\varphi_i;\psi_i$ and $\varphi_{i+1};\psi_i$ are shown below. The structural NMR ensemble of human insulin in solution (PDB id. 2mvc) is included as a solution structure of bovine insulin is currently not deposited in the PDB to the best knowledge of the authors. The crystal structure of bovine insulin (PDB id.: 2a3g) is included below. The secondary structure indications of the database are shown to ease comparison between the different figures. For the same reason the $(\varphi;\psi)$ angles (-66°;-41°), (-59°;-44°), (-71°;-18°) and (-49°;-26°) are shown.

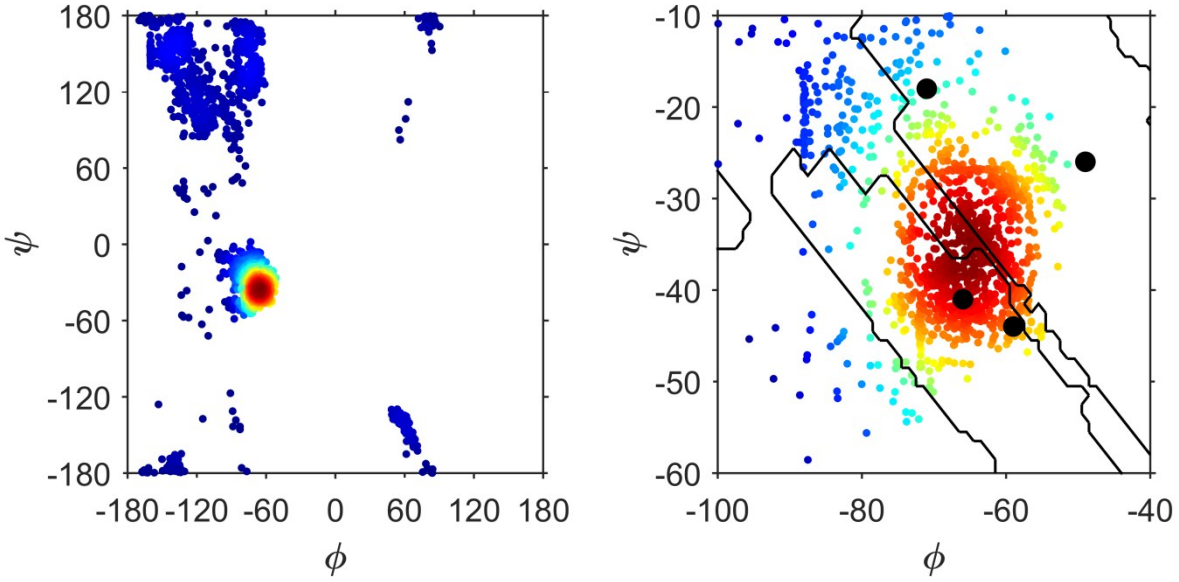


Figure S 20 $\varphi_i;\psi_i$ of the NMR solution structural ensembleof human insulin (PDB id. 2mvc).

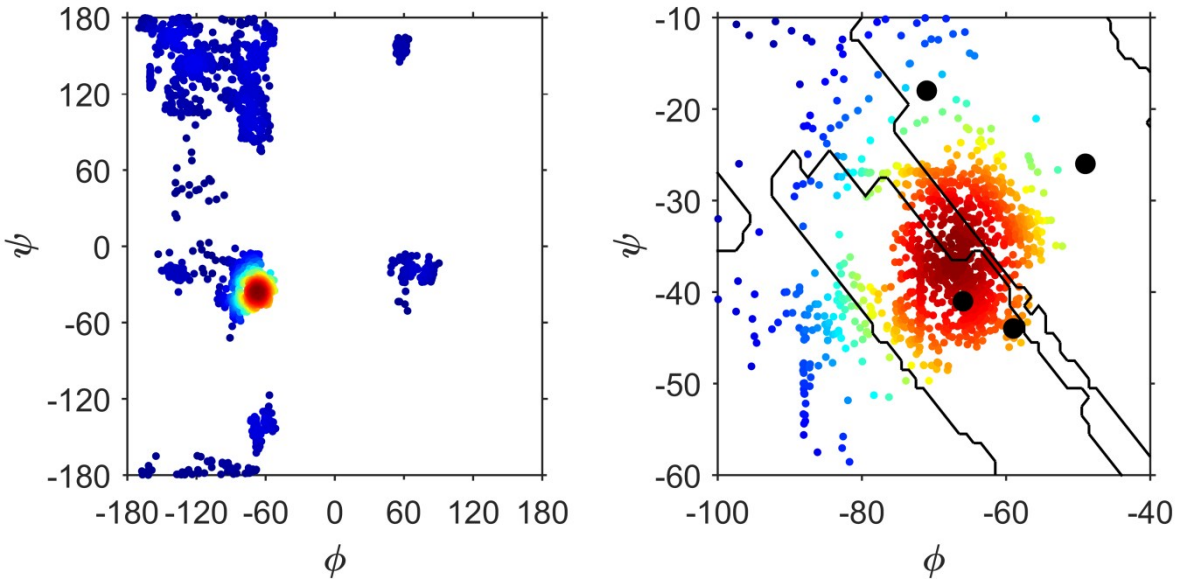


Figure S 21 $\varphi_{i+1};\psi_i$of the NMR solution structural ensembleof human insulin (PDB id. 2mvc).

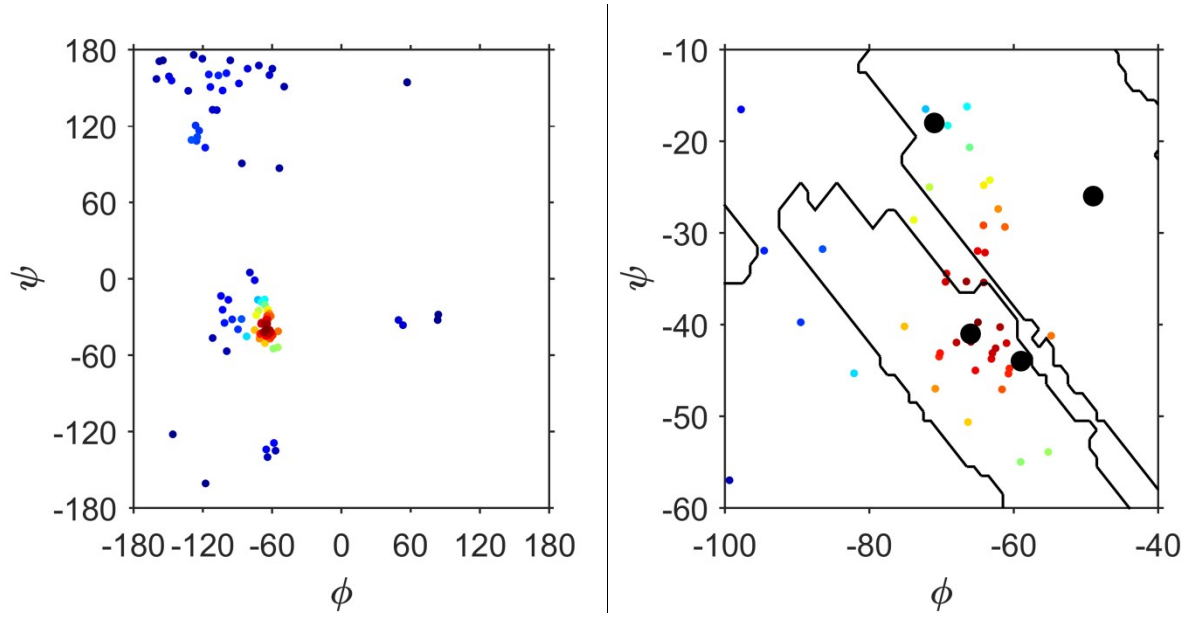


Figure S 22 φ_i;ψ_i of the crystal structure of bovine insulin (PDB id.2a3g).

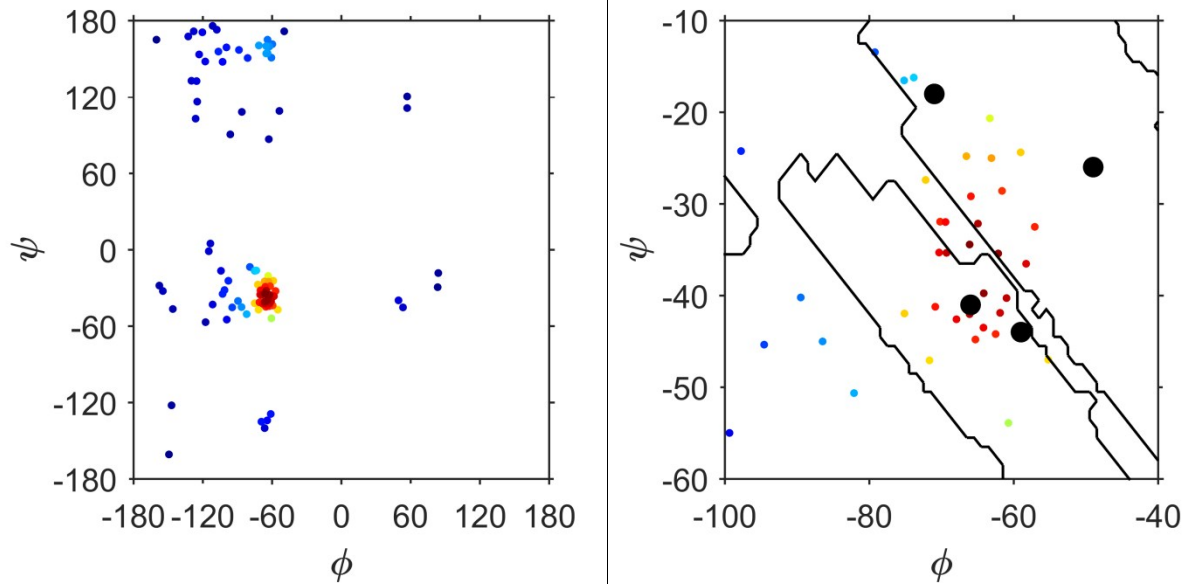


Figure S 23 φ_{i+1};ψ_i of the crystal structure of bovine insulin (PDB id.2a3g).

S10. Raman spectrum of PLA

The Raman spectrum of PLA is very comparable to that of AK21 but lacks the lower wavenumber region and is moved to the ESI for these reasons. Again, the database spectra describe the experimental patterns very well.

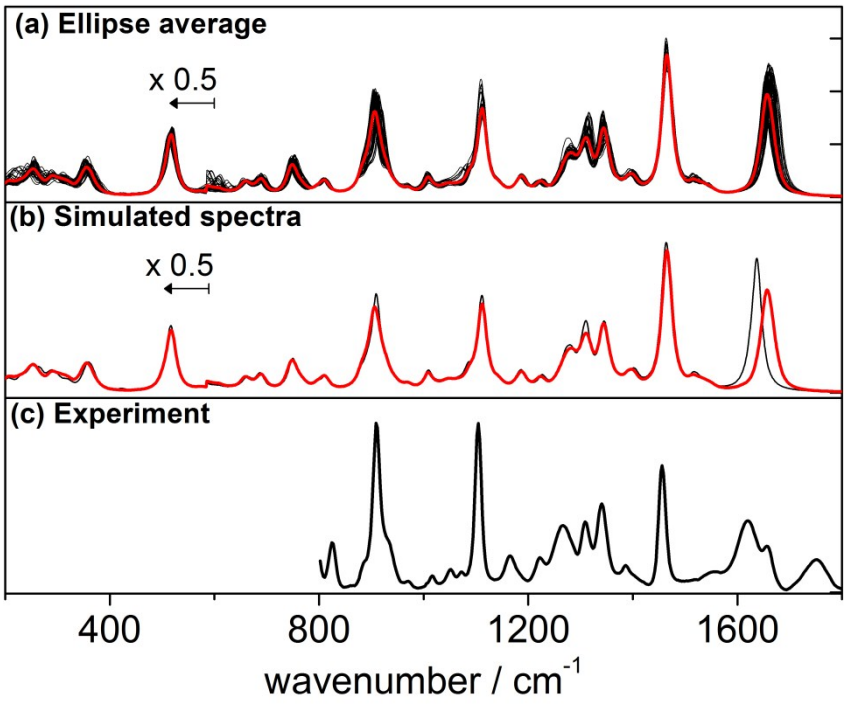


Figure S 24 Raman spectra of PLA: (a) spectra delineated by the ellipse in Figure S 9, (b, red) average spectrum of the ellipse spectra, (b, black) spectrum with the highest S_{fg}. compared to (c) the experimental spectrum.

S11. ESI References

1 D. A. Long, *The Raman Effect*, John Wiley & Sons, Ltd, Chichester, UK, 2002.

2 L. D. Barron, *Molecular light scattering and optical activity*, Cambridge University Press, Cambridge, second edi., 2004.

3 E. Debie, E. De Gussem, R. K. Dukor, W. Herrebout, L. A. Nafie and P. Bultinck, *ChemPhysChem*, 2011, **12**, 1542–1549.

4 I. K. McDonald and J. M. Thornton, *J. Mol. Biol.*, 1994, **238**, 777–793.