This item is the archived peer-reviewed author-version of:

Solvation and the secondary structure of a proline-containing dipeptide: insights from VCD spectroscopy

Reference:

Vermeyen Tom, Merten Christian.- Solvation and the secondary structure of a proline-containing dipeptide: insights from VCD spectroscopy
Physical chemistry, chemical physics / Royal Society of Chemistry [London] - ISSN 1463-9076 - 22:27(2020), p. 15640-15648
Full text (Publisher's DOI): https://doi.org/10.1039/D0CP02283G
To cite this reference: https://hdl.handle.net/10067/1709380151162165141

uantwerpen.be
Institutional repository IRUA

Solvation and the secondary structure of a proline-containing dipeptide: Insights from VCD spectroscopy

Tom Vermeyen,[a,b] Christian Merten[a]

Received 00th January 20xx,
Accepted 00th January 20xx

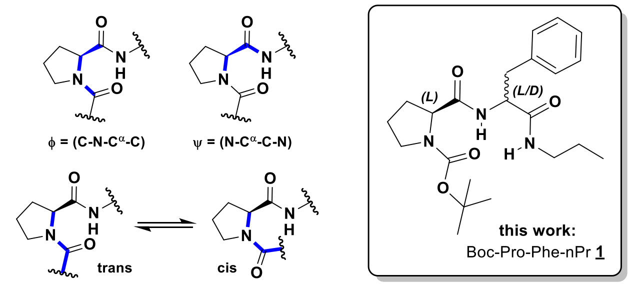
DOI: 10.1039/x0xx00000x

In this study we investigate the IR and VCD spectra of the diastereomeric dipeptides Boc-Pro-Phe-(n-propyl) **1** in chloroform-d_1 ($CDCl_3$) and the strongly hydrogen bonding solvent dimethylsulfoxide-d_6 (DMSO-d_6). From comparison of the experimental spectra, the amide II spectral region is identified as marker signature for the stereochemistry of the dipeptide: The homochiral LL-**1** features a (+/-)-pattern the amide II region of the VCD spectrum, while the amide II signature of the diastereomer LD-**1** is inverted. Computational analysis of the IR and VCD spectra of LL-**1** reveals that the experimentally observed amide II signature is characteristic for a β_I-turn structure of the peptide. Likewise, the inverted pattern found for LD-**1** arises from a β_{II}-turn structure of the dipeptide. Following a micro-solvation approach, the experimental spectra recorded in DMSO-d_6 are computationally well reproduced by considering only a single solvent molecule in a hydrogen bond with N-H groups. Considering a second solvent molecule, which would lead to a cleavage of intramolecular hydrogen bonds in **1**, is found to give a significantly worse match with the experiment. Hence, the detailed computational analysis of the spectra of LL- and LD-**1** recorded in DMSO-d_6 confirms that the intramolecular hydrogen bonding pattern, that stabilizes the β-turns and other conformations of LL- and LD-**1** in apolar solvents, remains intact. Our findings also show that it is essential to consider solvation explicitly in the analysis of the IR and VCD spectra of dipeptides in strongly hydrogen bonding solvents. As the solute-solvent interactions affect both conformational preferences and spectral signatures, it is also demonstrated that this inclusion of solvent molecules cannot be circumvented by applying fitting procedures to non-solvated structures.

Introduction

Unlike the other natural amino acids, the conformational space of the proline residue is restricted as the five-membered ring structure prevents rotation around the C^{a}-N bond. Consequently, proline can only adopt a very narrow range of ϕ-values ranging from -90° to -60° (cf. Scheme 1).[1] It is therefore well known as turn-inducing residue and as such found, for instance, in the $(i+1)^{th}$ position of β-turns.[2] Following Venkatachalam's pioneering conformational energy calculations,[3] the exact structure of a β-turn depends on the nature of the amino acid (polar/non-polar), its relative position in the sequence and its stereochemistry. Homochiral LL and heterochiral LD sequences, for instance, are known to favor β_I- and β_{II}-turn structures respectively.

Over the decades, many experimental and computational studies have been performed in order to deepen the understanding of secondary structure formation.[4-11] With regards to turn structures, these range from x-ray and NMR-studies on globular proteins to studies on small model peptides. Vibrational circular dichroism (VCD) spectroscopy, the chiroptical version of IR spectroscopy, has been used to study biomolecular structures since the 1980s.[12-16] Like for classical electronic CD, characteristic spectral signatures for certain classes of secondary structures have been derived and can be used to make qualitative statements about structural preferences.[12] In this respect, the knowledge on VCD signatures of proteins is well established.



Scheme 1. Important torsional angle definitions of the proline residue and the structure of the investigated dipeptide Boc-Pro-Phe-nPr **1**. Both diastereomeric pairs, LL/DD and LD/DL, were prepared.

We have recently studied a series of cyclic tetrapeptides of the general structure cyclo(Boc-L-Cys-L-Pro-*X*-L-Cys-OMe) by VCD spectroscopy and investigated the influence of hydrogen bonding organic solvents on their secondary structures.[17, 18] For all evaluated amino acids X in the 3-position, the aforementioned structural preferences for homo- and heterochiral sequences could be confirmed: The L-Ala version of the cyclic tetrapeptides, for instance, adopts a β_I-turn structure, while the D-Ala version prefers a β_{II}-turn structure. These secondary structure preferences give rise to characteristic VCD signatures in the amide II region, with β_I showing a (+/-) and β_{II} a (-/+) couplet. Interestingly, the VCD spectra also revealed a solvent dependence of the structure of the Gly-version of the peptide: In chloroform-d_1 the peptide shows a clear preference for β_{II}, strongly hydrogen bonding solvents stabilizes the β_I-turn structure. With regard to the spectra calculations it is important to note that it was not necessary to consider the solvent explicitly; the Gly-version of the peptide in the respective secondary

---

[a] Ruhr-Universität Bochum, Fakultät für Chemie und Biochemie, Organische Chemie II, Universitätsstraße 150, 44801 Bochum, Germany, E-mail: christian.merten@ruhr-uni-bochum.de, Web: www.mertenlab.de.
[b] University of Antwerp, Department of Chemistry, MolSpec group. Groenenborgerlaan 171, 2020 Antwerp, Belgium

Electronic Supplementary Information (ESI) available: Conformational analysis, optimized structures and further spectra plot. See DOI: 10.1039/x0xx00000x structure conformation already gave the experimentally observed spectral patterns.

In light of these findings, we decided to further investigate conformational preferences of small proline-containing peptides using VCD spectroscopy. In particular, we wanted to address two questions arising from our previous studies on the tetrapeptides: (1) Does a model system like Boc-Pro-Phe-nPr **1** (Scheme 1), which constitutes the smallest fragment necessary to form the hydrogen bonding network of a β-turn, already give the characteristic amide II signatures of β_{I} (+/-) to β_{II} (-/+) turns? (2) How does strong hydrogen bonding to the solvent influence the pattern? To this end, we prepared both diastereomeric pairs of **1** and recorded their IR and VCD spectra in chloroform-d_1 (which we consider herein as representative of non-polar solvents it typically does not alter conformational preferences) and DMSO-d_6 as strongly hydrogen bonding solvent. As we will show in this contribution, these model peptides allowed us to confirm that the observed amide II signatures are indeed characteristic for the β-turn structure. Furthermore, we show that solvation effects can be observed in the spectra of **1** and that explicit consideration of the solvent in the spectra calculations is necessary to obtain a good match between experiment and theory.

EXPERIMENTAL AND COMPUTATIONAL DETAILS

Materials. Starting materials, reactants and solvents for the synthesis were obtained from Sigma Aldrich at highest available purity and used without further purification. The deuterated solvents for the measurements were purchased from Eurisotop. DMSO-d_6 was stored over molecular sieve to minimize the water content.

Synthesis. The dipeptides **1** were prepared by coupling the corresponding phenylalanine n-propyl amide[19] with commercially available Boc-protected proline.[20]

IR and VCD spectroscopy. The IR and VCD spectra were recorded on a Bruker Vertex 70 equipped with a PMA 50 unit for polarization modulated measurements. Samples were held in BaF_2 cells with 210 µm path length. Concentrations were adjusted so that no IR bands of the sample were above 0.9 absorbance units. Final concentrations were 50 mM for all experiments except for LD/DL in DMSO-d_6, for which the concentrations were lowered to 35 mM due to low solubility. The IR spectra were accumulated for 32 scans, while the VCD spectra were recorded over a total measurement time of 8 hours each (34000 scans). Background correction was carried out by subtraction of the solvent spectrum respective to that of the racemic mixture recorded under identical conditions. The spectra recorded for DMSO-d_6 solutions are cut below 1100 cm^{-1} due to strong solvent absorbance.

Computational details. The conformational analysis was carried out by manually rotating around all the relevant dihedral angles, generating suitable starting structures. Geometry optimizations and spectra calculations were performed using Gaussian 09 Rev. E.01[21] at the B3LYP/6-31G(2d,p) level of theory. Solvation was accounted for implicitly in all calculations by using the IEFPCM[22-24] of the respective solvent. Additionally, in case of DMSO-d_6, the hydrogen bonding to the solvent was considered explicitly by placing solvent molecules at a S=O··H-N distance of about 2-2.5 Å in an arbitrarily chosen relative spatial orientation near the N-H groups of the peptide. Whenever identical conformers with different solvent orientations were found, the lowest energy conformer was selected while the others were discarded. Spectra were simulated by assigning a uniform Lorentzian band shape of 8 cm^{-1} half-width at half-height to the computed dipole and rotational strength. In order to account for errors due to the harmonic approximation, the computed frequencies were uniformly scaled by a factor of 0.976, which was visually determined to fit best with the experimentally observed spectra.

RESULTS AND DISCUSSION

Comparison of experimental spectra

The fingerprint region of the experimental IR and VCD spectra of both pairs of diastereomers of **1** are shown in Figure 1. From visual comparison, it can directly be noted that the IR spectra of the peptides change characteristically when changing the solvent from rather non-polar to polar, i.e. from $CDCl_3$ to DMSO-d_6. The amide I region (1750-1600 cm^{-1}) is characterized by C=O stretching vibrations and often regarded as the most sensitive pattern for secondary structure determination. In the present case, the exchange of the solvent leads to a clear change in band shape and thus suggests a structural change. Similarly, the amide II region of the NH-bending modes (1600-1450 cm^{-1}) changes characteristically from two distinct band maxima to a blue-shifted broad band. The amide III region of skeletal deformation modes (1320-1220 cm^{-1}) is not characteristic in the present case. Lastly it should be mentioned that the IR spectra of the diastereomers do not differ much and solely a minor change in line shape of the amide I and of the band at ~1400 cm^{-1} can be noted.

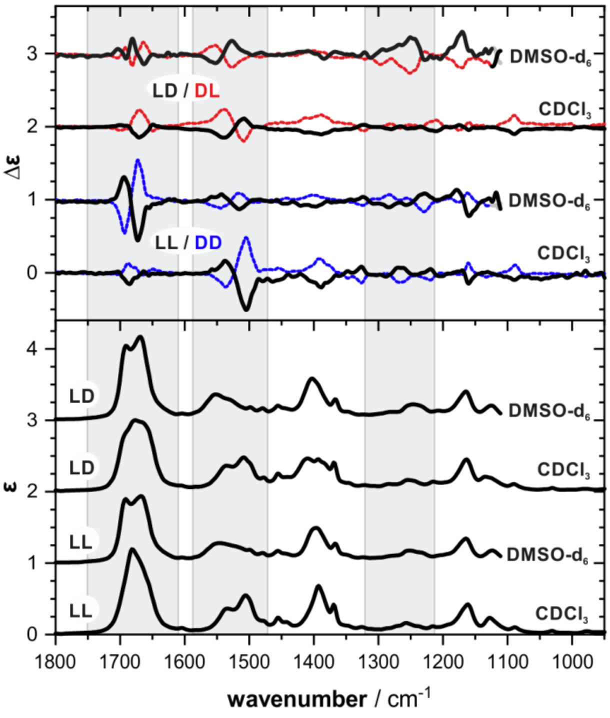


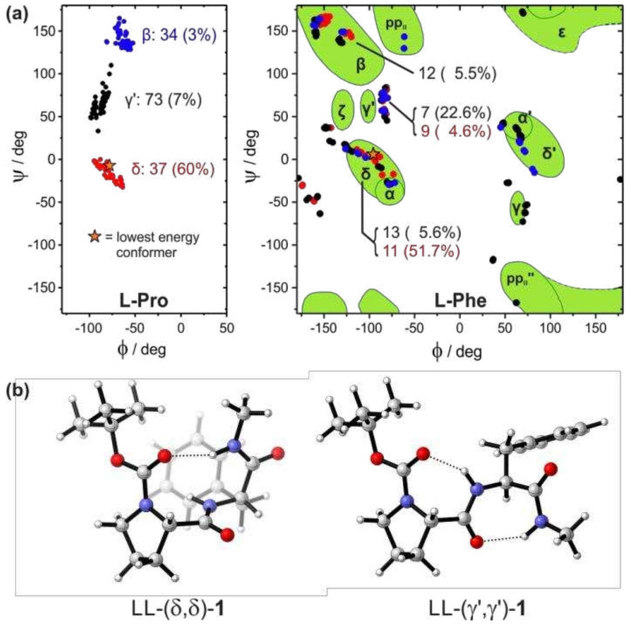
Figure 1. Experimental IR and VCD spectra of LL/DD-**1** (black/blue) and LD/DL-**1** (black/red) in CDCl$_3$ and DMSO-d$_6$. The grey shaded spectral ranges are the amide I-III regions. Molar absorptivity ε given in 10^3 M^{-1} cm^{-1} and differential molar absorptivity $\Delta\varepsilon$ given in 10^{-1} M^{-1} cm^{-1}. Experimental conditions: 210 µm path length and concentrations of 50 mM, except for LD/DL in DMSO-d$_6$, for which the concentrations were lowered to 35 mM due to low solubility.

In contrast to the IR spectra, the VCD spectra of both pairs of diastereomers differ significantly. For LL-**1** in CDCl$_3$, the signature of the amide I vibrations is weak and the amide II presents a strong (+/-)-couplet. In DMSO-d$_6$, however, the intensity distribution is reversed and the amide I shows a strong couplet while the intensity of the amide II (+/-)-couplet dropped significantly. Sign inversion is also observed for bands in the amide III region. For LD-**1**, the amide I and amide II are of comparable intensity in both solvents, but the amide III region gains significantly in intensity in DMSO-d$_6$. The difference in stereochemistry also becomes obvious as there is a sign change in the amide II region: A (-/+)-pattern is found for LD-**1** in both solvents. Such sign change has also been observed in our previous studies on cyclic tetrapeptides[17, 18] and could be correlated with a transition from β_{I} (+/-) to β_{II} (-/+) turn structure. Summarizing this comparison of the experimental spectra, two main conclusions can be drawn. First, the peptides are barely distinguishable on the level of IR spectroscopy but easily by their VCD signatures. Secondly, solvation with DMSO-d$_6$ obviously changes the conformational preferences of both peptides in a characteristic way. In the following sections, the computational analysis of LL-**1** is discussed in detail to show that the spectral signatures can be directly correlated with the conformational preferences of the peptide.

Analysis of the spectrum of LL-1 recorded in CDCl$_3$

The conformational space of dipeptide **1** is complex and it can theoretically adopt many conformations. Consequently, in order to correlate the changes in the VCD spectral signatures with structural preferences, a comprehensive conformational search has been carried out first. Instead of using molecular dynamics or Monte Carlo approaches, we systematically explored the conformational space by manually generating starting structures for geometry optimizations. We evaluated the peptide backbone angles, i.e. various ϕ/ψ-angles for the two amino acid residues (cf. Scheme 1), the proline ring puckering and the orientation of the benzyl side group. The trans/cis-isomerization of the Boc-Pro bond has been evaluated for several conformer families. In order to simplify the analysis, the n-propyl group was truncated to a methyl group. Furthermore, despite CDCl$_3$ being a weak hydrogen bond donor, we considered the solvent only implicitly by using a continuum solvation model (IEFPCM).

Following this procedure, we obtained a set of over 140 unique conformers of LL-**1**. In order to visualize the sampled conformational space, each of these structures is represented by a dot in the Ramachandran plots in Figure 2a. The numbers given in the plot summarize the populations accumulated in selected regions (cf. Table S1 for detailed numbers on all regions). The right panel of Figure 2a also indicates the nomenclature for the regions that we use throughout this study.[25, 26] The rigidity of the Pro residue is reflected in the small range of ϕ-angles: It can adopt only the δ-, γ'-, and pp$_{\text{II}}$-conformations, while the conformational weight is mainly distributed over the first two regions. Despite the higher conformational flexibility of the L-Phe residue, the lowest energy structures of LL-**1** are localized in the (δ,δ)- and (γ',γ')-regions (Fig. 2b). All other areas are not significantly populated. The (δ,δ)-conformation resembles a β_{I}-turn structure with an $i \rightarrow i+3$ intramolecular hydrogen bond (the Boc-group is considered residue *i*, the n-propyl/methyl amine residue *i+3*). In the (γ',γ')-conformation, LL-**1** features two intramolecular hydrogen bonds ($i \rightarrow i+2$ and $i+1 \rightarrow i+3$), which are commonly recognized as γ-turn structures, that stabilize a flat geometry of the backbone. As shown in Figure 2a, the two regions of the (δ,δ) and (γ'γ')-conformers together also account for almost 75 % of the Boltzmann population; the two conformers shown in Figure 2b contribute individually with 26 and 14 %. The conformer family corresponding to a β_{II}-turn structure, that is, the (δ,δ')-family, is not populated and its lowest energy representative is 2.5 kcal/mol higher in energy than the global minimum.



LL-(δ,δ)-**1** LL-(γ',γ')-**1**

Figure 2. Conformational analysis of LL-**1**. (a) Ramachandran plots showing the conformational distribution, the number of conformers and in parenthesis the contribution of their conformer families in percentage according to ΔE_{ZPC}; (b) Lowest energy conformation of LL-**1** in the (δ,δ)- and (γ',γ')-families (populations of 26.0 respectively 13.7 %).

The computational analysis of the conformational space of LL-**1** shows the same preference for a β_{I}-turn structure that we have observed previously for the cyclic tetrapeptides.[17, 18] In the next step, we therefore simulated the IR and VCD spectra of LL-**1** in order to confirm the predicted conformational distribution. To this end, the single-conformer spectra were scaled by their respective Boltzmann weights as determined based on the relative zero-point corrected energies ΔE_{ZPC} and co-added to obtain the conformationally averaged spectra shown in Figure 3. The almost perfect match between the computed and experimental IR spectrum can be noted immediately. Solely the proline C=O stretching vibration, that does not participate in an intramolecular hydrogen bond in the (δ,δ)-conformers, is predicted at a higher frequency than observed in the experiment so that it appears as a rather isolated band instead of merged with the amide I vibrations. In order to further emphasize the very good match between experiment and theoretical prediction, the computed VCD spectra of LL-**1** are compared with the experimental spectra of DD-**1**. The almost perfect mirror-image relation shows that not only the amide I and amide II regions are well reproduced, but also most of the small signatures below 1450 cm^{-1} are predicted correctly.

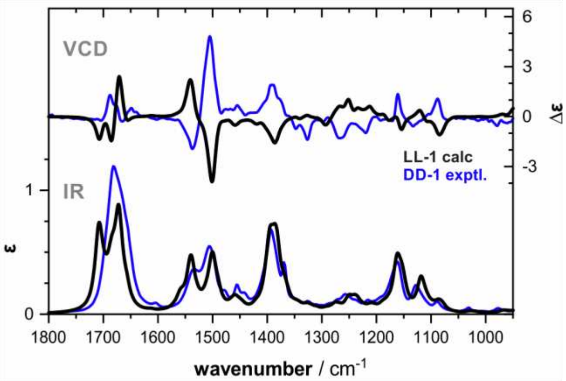
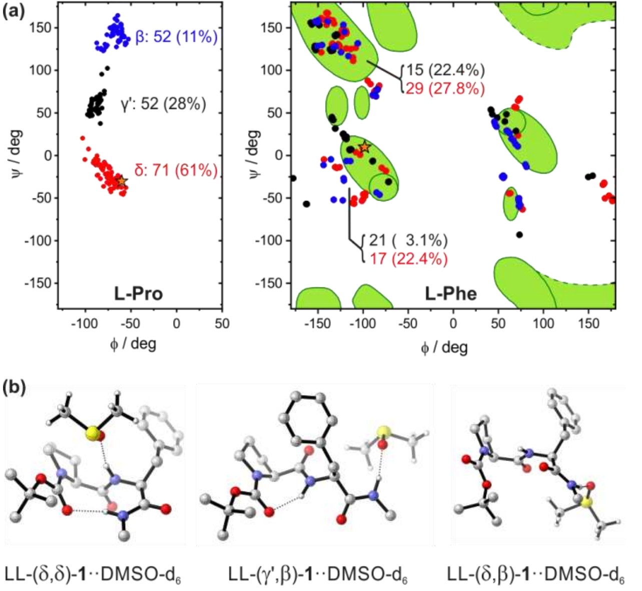


Figure 3. The computed IR and VCD spectra of LL-**1** overlapped with the experimental spectra of DD-**1** recorded in CDCl$_3$. Molar absorptivity ε given in 10^3 M^{-1} cm^{-1} and differential molar absorptivity Δε given in 10^{-2} M^{-1} cm^{-1}.

The dipeptide LL-1 in DMSO-d$_6$: Intra- vs intermolecular hydrogen bonding

The very good match between experimental and computed spectra confirms the predicted conformational preferences of LL-**1** in chloroform-d$_1$ respectively in a non-hydrogen bonding solvent. Hydrogen bonding of the amides' N-H groups to DMSO-d$_6$, however, is expected to play a major role in determining the conformational preferences of the peptide and the solvent is likely to compete with intramolecular interactions as well. Therefore, when analyzing spectra recorded in this or other strongly hydrogen bonding solvents, it is usually necessary to include the solute-solvent interaction explicitly in the conformational analysis and subsequent spectra calculations.[27-31] For LL-**1**, it also has to be taken into account that solute-solvent interactions may occur with either one of the N-H groups at a time or with both of them simultaneously (i.e. two different molecules of DMSO-d$_6$ or one in a bifurcated interaction).[27] Hence, the conformational analysis becomes quite complex when all these possibilities are explored.

Nonetheless, based on the set of conformers discussed above, we began to compute mono-solvated structures of LL-**1** by explicitly placing one molecule of DMSO-d$_6$ near the N-H groups. In this process, we tried to preserve the hydrogen bonding network in each of the conformers. For instance, in order to keep the intramolecular hydrogen bond characteristic for the β$_I$-turn in the (δ,δ)-conformations (lowest energy conformer family; cf. Fig 2b), we placed the DMSO-d$_6$ molecule only near the free NH-group of the Phe residue. Consequently, for some conformer families this meant that two different mono-solvated structures could be built as both of their NH-groups were not participating in intramolecular hydrogen bonds and thus solvent accessible. In turn, this also meant that DMSO-solvated structures of (γ',γ')-structures of LL-**1**, the second lowest energy conformer family of the non-solvated form, could not be built as this would have required to break the intramolecular hydrogen bonding network. In addition, with the (pp$_{II}$,α)- and (δ,α)-conformer classes it was possible to build structures with both NH-groups interacting with the same molecule of DMSO-d$_6$, i.e. with DMSO-d$_6$ being bound in a bifurcated hydrogen bonding pattern.



LL-(δ,δ)-**1**··DMSO-d$_6$ LL-(γ',β)-**1**··DMSO-d$_6$ LL-(δ,β)-**1**··DMSO-d$_6$

Figure 4. Conformational analysis of LL-**1**··DMSO-d$_6$. (a) Ramachandran plot showing the ϕ/ψ-distribution of the computed mono-solvated structures and the accumulated Boltzmann weights according to ΔE$_{ZPC}$. (b) The three lowest energy structures of LL-**1**··DMSO-d$_6$. Non-polar hydrogen atoms, except those of the CH$_3$-groups in DMSO, were omitted for clarity.

Taking into account all these aspects, we obtained almost 170 conformers of LL-**1**··DMSO-d$_6$. As visualized in Figure 4a, the increased conformational flexibility of mono-solvated LL-**1** resulted in a more widespread distribution over the ϕ/ψ-backbone angle space. Especially conformers in the δ- and β-Phe regions are spread out over a larger range, as solvation of either NH-group and trans/cis isomerization increased the number of structures. The large number of conformers on the right of the δ-Phe region is mostly due to conformers with bifurcated hydrogen bonding. Figure 4b shows the three lowest energy conformers of LL-**1**··DMSO-d$_6$, which are representatives of the (δ,δ)-, (γ',β)- and (δ,β)-conformer families. These three conformers contribute individually about 11.2 % (δ,δ), 10.8 % (γ',β) and 7 % (δ,β) to the overall conformational distribution, while their families have cumulated Boltzmann weights of 22 % (δ,δ), 22 % (γ',β) and 28 % (δ,β). Without considering the explicit hydrogen bonding interactions to DMSO-d$_6$, that is, in the analysis of the data recorded in chloroform-d$_1$, the contributions of the (γ',β)- and (δ,β)-families were negligible with 5.5 and 3.7 %. Explicit solvation with one molecule of DMSO-d$_6$ therefore clearly introduces a change in conformational preferences, which leads away from the β$_I$-turn conformer family and towards structures with more solvent-accessible N-H groups.

Notably, the (δ,β)-conformer in Figure 4b is a cis-conformer with the corresponding trans-isomer being about 0.55 kcal/mol higher in energy. This preference is, however, not associated with solvation, as it can also be observed in the non-solvated structures. Rather it arises from a more preferential, alternating alignment of the C=O dipoles that is achieved in the cis-conformation. Especially due to the shift of the conformational preferences towards the (δ,β)-family, the computed overall ratio of trans/cis-isomers for LL-**1**··DMSO-d$_6$ is ~60:40 (as opposed to 92:8 in the non-solvated case).

Based on the computed conformational preferences of the explicitly mono-solvated structures of LL-**1** in DMSO-d$_6$, we simulated the IR and VCD spectra and compare them to the experimental spectra of DD-**1** in Figure 5. The overlap of the IR spectra again reveals a generally very good match between experimental and theoretical spectra. Minor deviations in the IR band shape in the amide II region and of the IR signature in the range 1400-1350 cm^{-1} can be noted, which we attribute to slight frequency deviations and differences in band width between the simulated and experimental spectra. Likewise, the predicted VCD spectral pattern again resembles the experimental signatures over the full investigated range. In this respect, it is again very noteworthy that not only the amide I/II regions but even the small VCD bands in the spectral range below 1450 cm^{-1} are nicely reproduced.

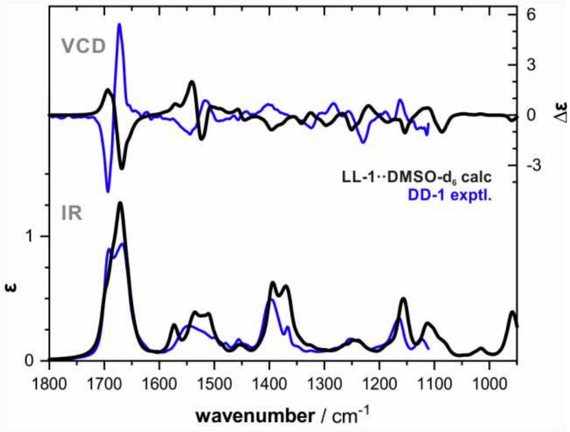


Figure 5. Overlap of the computed IR and VCD spectra of the mono-solvated forms LL-**1**··DMSO-d$_6$ with the experimental IR and VCD spectra of DD-**1** recorded in DMSO-d$_6$. Molar absorptivity ε given in 10^3 M^{-1} cm^{-1} and differential molar absorptivity Δε given in 10^{-2} M^{-1} cm^{-1}.

As preliminary conclusion, it can be stated that the predicted conformational preferences for the mono-solvated forms based on the ΔE_{ZPC} energies appear to capture the experimental conformational distribution quite well. However, as discussed above, having two N-H groups makes LL-**1** capable of simultaneously forming hydrogen bonds to two different DMSO-d$_6$ molecules. Following the approach of not breaking intramolecular interactions during the preparation of solvated structures of LL-**1**, we could only solvate those conformers which had two accessible NH-groups. As a consequence, many conformer families including, for instance, the entire class of β_I-turn structures could not be considered for twofold solvation. In the end, only the twofold-solvated (δ,β)-, (δ,γ')-, (pp$_{II}$,β)- and (pp$_{II}$,γ')-conformations could be built. The direct comparison of the experimental spectra with the computed IR and VCD spectra of these few conformers showed no match in the amide II region (all positive instead of (+/-)-pattern; Fig. S1). In fact, also the other parts of the spectra, except for the amide I pattern, were significantly different from the experimental spectra. Therefore, it can be concluded that sole twofold solvation and neglect of many intramolecular hydrogen bonded conformations in fact yields an incorrect picture of the conformational preferences of LL-**1**.

The effect of DMSO-d$_6$ solvation on conformer family spectra

Tackling effects of solute-solvent hydrogen bonding of molecules with increasing conformational flexibility and, in particular, with several competing intramolecular hydrogen bonding interactions is obviously time-consuming and challenging. Searching for a way to simplify the analysis, it is tempting to bring together two approaches yet discussed separately in literature: (1) For Raman Optical Activity (ROA) spectra of proteins, Mensch and Johannessen have recently shown that a database of computed spectra of model structures with different secondary structures can be used to predict predominant secondary structures.[32, 33] (2) Aiming for an improved match between experimental and theoretical spectra, Buma et al. have recently introduced a fitting procedure[34] to adjust conformer weights in the VCD spectral analysis of small molecules. One could imagine a fitting procedure that takes representative spectra, i.e. those of the conformer families of LL-**1** (not LL-**1**··DMSO-d$_6$!) that are easily obtained without sampling the solvent, and determines the optimum combination (conformer family weights) that give the best fit with the experimental spectra. If successful, this could be a pragmatic approach to circumvent the time-consuming consideration of explicit solvation.

In order to challenge this idea, Figure 6a shows the simulated VCD spectra of the four main conformer families with and without consideration of explicit solvation with DMSO-d$_6$. The spectra of the β_I-turn family, i.e. of the (δ,δ)-conformers, both feature a clear (+/-) signature for the amide II region which matches with the experimentally observed patterns. This confirms that the amide II signature is indeed a characteristic of the β_I-turn. The computed amide I region of the (δ,δ)-family, however, shows the opposite pattern to the one observed in chloroform-d$_1$ and also does not match the complex pattern recorded for the DMSO-d$_6$ solution of LL-**1**. This clearly shows that the contributions of the other conformer families to the amide I and II regions cannot be neglected in the spectral analysis. The (γ',γ')-conformer family, for instance, is the main contributor to the (+/-)-signature of the amide I region in the simulated VCD spectrum of monomeric LL-**1**.

When comparing the spectra within the same conformer family with and without explicit solvation, the spectral signatures found below 1400 cm^{-1} are rather unaffected by the solute-solvent hydrogen bond. The main differences are found in the amide I and II regions, although it must be noted that the differences in the (δ,δ)-conformer families' spectra are small compared to the changes found in the spectra of the other families. Nonetheless, given that the amide I and II regions are obviously the most conformer-sensitive bands in the VCD spectrum of LL-**1**, a fitting procedure as the proposed one is little likely to work. To demonstrate this, we simulated the VCD spectrum of LL-**1** in DMSO-d_6 using the (δ,δ), (γ',β) and (δ,β)-family spectra of the "non-solvated" structures and the cumulated Boltzmann weights of their solvated analogues (cf. Figure 6b). The resulting VCD spectrum clearly does not resemble the spectrum obtained using the solvated conformers. Hence, in the future, new routines will have to be developed that allow the (guided) automatic placement of solvent molecules or a reliable a priori prediction of important conformer classes.

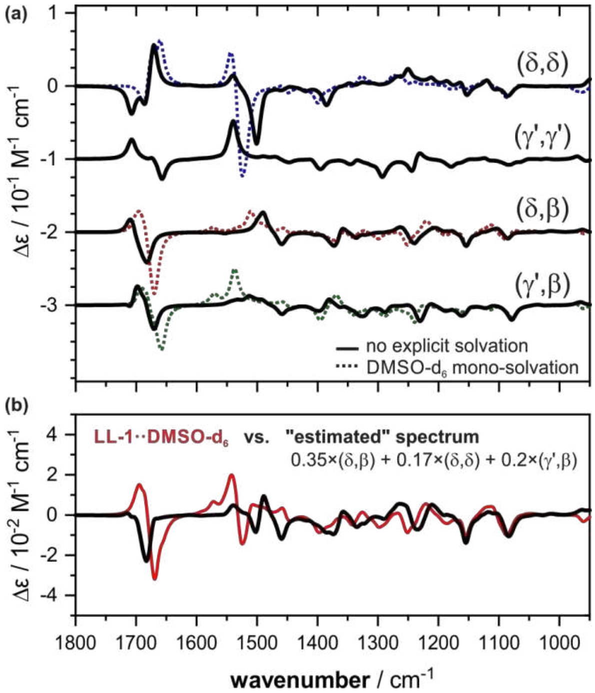


Figure 6. Conformer family spectra of LL-**1**. (a) The VCD spectra of the (δ,δ)-, (γ',γ')-, (γ',β) and (δ,β)-family spectra of LL-**1** with and without explicit consideration of solvation one molecule of DMSO-d_6. Note: The (γ',γ')-conformers cannot form hydrogen bonds with DMSO-d_6. (b) The "estimated" VCD spectrum is obtained from the family spectra of the "non-solvated" conformer families as shown in (a) but using the corresponding Boltzmann weights of the LL-**1**··DMSO-d_6 families. For comparison, the full spectrum of LL-**1**··DMSO-d_6 is shown as well.

When discussing the computation of conformer ratios, it must be pointed out that the above discussion of the conformational preferences was based on relative zero-point corrected energies, ΔE_{ZPC}, and the corresponding Boltzmann populations. Using the Boltzmann weights determined from relative Gibbs Free Energies ΔG_{298K} also gives IR and VCD spectra for LL-**1** that match well with the experimental spectra recorded in chloroform-d_1. The only noteworthy difference is found for the highest frequency band of the amide I region, which changes sign in the VCD spectrum. This and a few other minor differences are due to differences in the overall contributions of the conformer families between for ΔE_{ZPC} and ΔG_{298K}.. This is easily confirmed by comparing the IR and VCD spectra of the conformer families themselves, which are almost independent of the relative energies used to compute the Boltzmann weighing factors (cf. Fig. S2; Tab. S1 and S2). For LL-**1**··DMSO-d_6, however, spectra based on plain ΔG_{298K} do not reproduce the VCD signature in the amide II region, where only positive bands are predicted (cf. Fig. S3). The underlying drastic change in conformer weights is likely due to the occurrence of many low-frequency intramolecular vibrational motions introduced by the explicit inclusion of the solvent, which affect the computation of the entropic contribution to ΔG_{298K}. Following Cramer and Trular,[35] the entropic contribution can be corrected by raising all vibrational frequencies lower than 100 cm^{-1} to this threshold value and re-computing the partition function for the entropy. This procedure gives very similar conformer weights as ΔE_{ZPC}-energies (cf. population summary in Tab. S1), thus confirming our hypothesis. As we have also observed a generally better fit of the ΔE_{ZPC} for other examples in the past,[27-31] it seems justified to generally recommend the use of ΔE_{ZPC} instead of ΔG_{298K} for the simulation of IR and VCD spectra of micro-solvated clusters. Lastly, we also note that we briefly evaluated the effect of dispersion corrections (GD3BJ)[36] by re-optimizing the 14 lowest energy conformers of LL-**1**··DMSO-d_6. The resulting relative energies and spectra are shown in the supporting information. Similar to our observations in previous studies on microsolvation,[30, 37] the inclusion of GD3BJ caused a significant shift in the conformational preferences, making the (δ,δ)-conformer class the most dominant (>90%) and removing almost any contribution by the (δ,β) and (γ',β)-families. Consequently, as the IR and VCD spectra of the (δ,δ)-conformers looked essentially as those obtained without dispersion correction (cf. Figure 6), the final computed spectra did not match with the experiment.

The spectra of LD-1

Guided by the results obtained for LL-**1**, we also computed the IR and VCD spectra of LD-**1** and LD-**1**··DMSO-d_6. As the stereochemistry at the C$^{\alpha}$ of D-phenyl alanine is inverted, it is noted that the residue can adopt different ϕ/ψ-backbone angles. They are, however, generally related to the definitions of preferred conformers of L-Phe by a simple 180° rotation of the Ramachandran plot (cf. Fig. S5; note that we also rotate the nomenclature from Fig. 2a).[38] Our conformational search for LD-**1** gave over 100 conformers of which the lowest energy structures were found to be from the LD-(pp$_{II}$,^{D}δ)-**1** and LD-(δ,^{D}γ')-**1** conformer families (Fig. 7a). In the (pp$_{II}$,^{D}δ)-conformation, LD-**1** adopts a β$_{II}$-turn geometry and is thus the structural analogue to β$_{I}$-turn structure LL-(δ,δ)-**1**. The (δ,^{D}γ')-conformation of LD-**1** features two intramolecular hydrogen bonds (γ-turns) similar to LL-(γ',γ')-**1**, but the inverted stereochemistry of D-Phe results in an overall bended conformation. We also obtained almost 110 conformers of LD-**1**··DMSO-d_6 with the dominating conformer family being again the β$_{II}$-turn structure (pp$_{II}$,^{D}δ) with a population of 78 %.

The VCD spectra predicted based on the sets of LD-**1** and LD-**1**··DMSO-d_6 conformers are shown in Figure 7b (the IR spectra are presented in the supporting information.). Again a good match between experiment and theory is observed: Both the VCD pattern and also the relative intensity ratios between amide I and II regions are well reproduced. The analysis of the conformer family spectra shows that the amide II region of the (pp$_{II}$,^{D}δ)-conformers of LD-**1**, which are those in a β$_{II}$-turn, features a (-/+)-pattern. Hence, also for LD-**1** we can confirm the correlation of β-turn structure to characteristic amide II signatures.

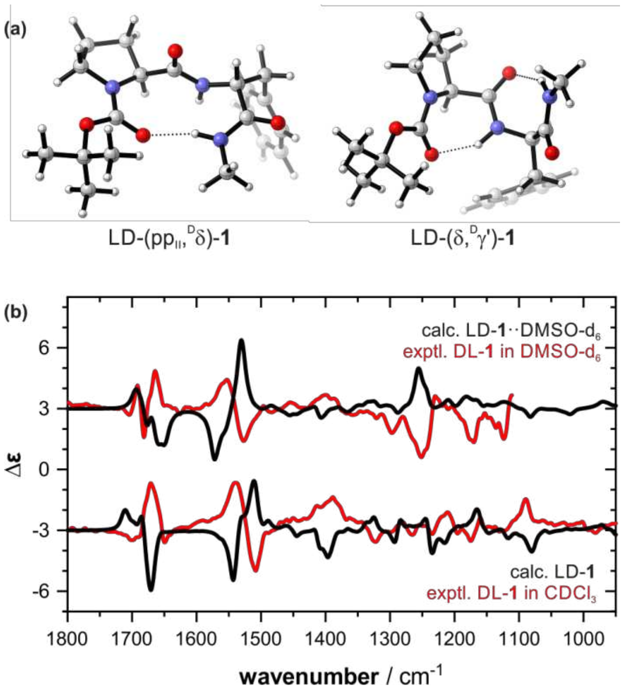


Figure 7. Analysis of the VCD spectra of LD-**1**. (a) Lowest energy conformations of LD-**1**. (b) The experimental VCD spectra of DL-**1** recorded in CDCl$_3$ and DMSO-d$_6$ compared to the computed spectra of LD-**1** and LD-**1**··DMSO-d$_6$. The differential molar absorptivity Δε given in 10^{-2} M^{-1} cm^{-1}.

Conclusions and Outlook

In this study, we investigated the IR and VCD spectra of the two diastereomeric dipeptides LL- and LD-**1** in CDCl$_3$ and DMSO-d$_6$. While the experimental IR spectra are found to be very similar when comparing the diastereomers in the same solvent, the VCD spectra show clear marker bands characteristic for the peptides' stereochemistry. Furthermore, differences in the IR spectral patterns of the same peptide in the two solvents were rather small, while the VCD spectra again revealed significant solvent dependent changes.

A comprehensive conformational analysis provided a solid basis for the prediction of theoretical spectra of LL-**1** (and likewise for LD-**1**) which almost perfectly matched with the experimental spectra recorded in CDCl$_3$. The computational results indicate two types of secondary structures as dominant conformer families, namely the (δ,δ)- and (γ',γ')-conformers. Characteristic spectral features could be identified that confirmed their contributions to the spectra. In particular, the β$_I$-turn structures (the (δ,δ)-conformers) are found to give rise to a characteristic (+/-)-pattern amide II for LL-**1** (and likewise a (-/+)-pattern for the β$_{II}$-turn structure of LD-**1**).

Based on the conformational preferences established in the first part of the spectral analysis, we considered explicit solvation with DMSO-d$_6$ in order to analyze the experimental data obtained for the hydrogen bond acceptor solvent. Considering only one solvent molecule and preserving intramolecular hydrogen bonds, the (δ,δ)-, (γ',β)- and (δ,β)-families were identified as the major conformers of LL-**1**··DMSO-d$_6$. Comparison of the computed IR and VCD spectra for the mono-solvated structures with the experimental spectra showed a very good match and thus confirmed the validity of the predicted conformational preferences. The computations also showed that the VCD signature of the β$_I$-turn in LL-**1** (and the β$_{II}$-turn in LD-**1**) is not affected by explicit solvation. Subsequently, explicit solvation of LL-**1** with two molecules of DMSO-d$_6$, that means having both N-H groups participating in intermolecular instead of intramolecular hydrogen bonds, was shown to not reproduce the experiment.

In conclusion, these observations confirm that intramolecular interactions in these simple peptides are already strong enough to stabilize secondary structure elements even in a strongly competing solvent such as DMSO-d$_6$. They agree with the observations made for larger cyclic peptides before.[17, 18] From the spectroscopic perspective it is additionally noted that, while the VCD patterns in the amide I and II regions of the β-turn conformers qualitatively not strongly affected, intensities and exact band positions in the spectra of most conformer families changed significantly upon explicit solvation. These two aspects taken together have strong general implications for the spectral analysis of peptides in strongly hydrogen bonding solvents. As underlined by our simulation experiment, the solvent must be considered explicitly in order to predict both the correct conformational distribution and the correct spectral signatures. Neither can be obtained by a simple fitting of monomeric spectra to the experiment in DMSO-d$_6$. This will require new approaches for conformational sampling which take into account the solvent environment and which go beyond explicit manual placement of single solvent molecules.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-2033; project no. 390677874) and the DFG's Heisenberg programme (ME 4267/5-1; project no. 418661145). Financial support by the Fund for Scientific Research-Flanders (FWO-Vlaanderen; grant number 1160419N) is acknowledged.

Notes and references

1. P. Chakrabarti and D. Pal, *Progress in Biophysics and Molecular Biology*, 2001, **76**, 1-102.
2. G. D. Rose, L. M. Glerasch and J. A. Smith, *Turns in Peptides and Proteins* in *Advances in Protein Chemistry*,

eds. C. B. Anfinsen, J. T. Edsall and F. M. Richards, Academic Press1985, vol. 37, pp. 1-109.

3. C. M. Venkatachalam, *Biopolymers*, 1968, **6**, 1425-1436.
4. Y. Yang, J. Gao, J. Wang, R. Heffernan, J. Hanson, K. Paliwal and Y. Zhou, *Briefings in Bioinformatics*, 2016, **19**, 482-494.
5. B. Rost, *Journal of Structural Biology*, 2001, **134**, 204-218.
6. S. P. Mielke and V. V. Krishnan, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2009, **54**, 141-165.
7. Y. Shi, *Cell*, 2014, **159**, 995-1014.
8. B. W. Matthews, *Annu. Rev. Phys. Chem.*, 1976, **27**, 493-493.
9. P. R. L. Markwick, T. Malliavin and M. Nilges, *PLOS Computational Biology*, 2008, **4**, e1000168.
10. A. Cavalli, X. Salvatella, C. M. Dobson and M. Vendruscolo, *PNAS*, 2007, **104**, 9615-9620.
11. P. J. Barrett, J. Chen, M.-K. Cho, J.-H. Kim, Z. Lu, S. Mathew, D. Peng, Y. Song, W. D. Van Horn, T. Zhuang, F. D. Sönnichsen and C. R. Sanders, *Biochemistry*, 2013, **52**, 1303-1320.
12. T. A. Keiderling, *Chem. Rev.*, 2020, **120**, 3381-3419.
13. T. A. Keiderling, *Curr. Op. Chem. Bio.*, 2002, **6**, 682-688.
14. T. B. Freedman, L. A. Nafie and T. A. Keiderling, *Biopolymers*, 1995, **37**, 265-279.
15. T. A. Keiderling, *Vibrational Circular Dichroism Spectroscopy of Peptides and Proteins* in *Circular Dichroism. Principles and Applications*, eds. K. Nakanishi, N. Berova and R. W. Woody, VCH Publishers Inc. and Weinheim and New York and Cambridge1994, pp. 497-522.
16. R. D. Singh and T. A. Keiderling, *Biopolymers*, 1981, **20**, 237-240.
17. C. Merten, F. Li, K. Bravo-Rodriguez, E. Sanchez-Garcia, Y. Xu and W. Sander, *Phys. Chem. Chem. Phys.*, 2014, **16**, 5627-5633.
18. N. Berger, F. Li, B. Mallick, J. T. Brüggemann, W. Sander and C. Merten, *Biopolymers*, 2017, **107**, 28-34.
19. M. C. Holland, J. B. Metternich, C. Daniliuc, W. B. Schweizer and R. Gilmour, *Chem. Eur. J.*, 2015, **21**, 10031-10038.
20. P. Fatás, A. I. Jiménez, M. I. Calaza and C. Cativiela, *Org. Biomol. Chem.*, 2012, **10**, 640-651.
21. *Gaussian 09, Rev E.01*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Wallingford CT, USA, 2013
22. B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel and P. J. Stephens, *J. Phys. Chem. A*, 2002, **106**, 6102-6113.
23. J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999-3094.
24. B. Mennucci, C. Cappelli, R. Cammi and J. Tomasi, *Chirality*, 2011, **23**, 717-729.
25. A. Hollingsworth Scott and P. A. Karplus, in *BioMolecular Concepts*2010, vol. 1, p. 271.
26. O. Carugo and K. Djinovic-Carugo, *Acta Crystallographica Section D*, 2013, **69**, 1333-1341.
27. L. Weirich, J. Magalhaes de Oliveira and C. Merten, *Phys. Chem. Chem. Phys.*, 2020, **22**, 1525-1533.
28. L. Weirich and C. Merten, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13494-13503.
29. K. Bünnemann, C. H. Pollok and C. Merten, *J. Phys. Chem. B*, 2018, **122**, 8056-8064.
30. K. Bünnemann and C. Merten, *Phys. Chem. Chem. Phys.*, 2017, **19**, 18948-18956.
31. K. Bünnemann and C. Merten, *J. Phys. Chem. B*, 2016, **120**, 9434-9442.
32. C. Mensch, P. Bultinck and C. Johannessen, *ACS Omega*, 2018, **3**, 12944-12955.
33. C. Mensch, L. D. Barron and C. Johannessen, *Phys. Chem. Chem. Phys.*, 2016, **18**, 31757-31768.
34. M. A. J. Koenis, Y. Xia, S. R. Domingos, L. Visscher, W. J. Buma and V. P. Nicu, *Chemical Science*, 2019, **10**, 7680-7689.
35. R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2011, **115**, 14556-14562.
36. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
37. N. M. Kreienborg and C. Merten, *Chem. Eur. J.*, 2018, **24**, 17948-17954.
38. C.-L. Towse, G. Hopping, I. Vulovic and V. Daggett, *Protein Engineering, Design and Selection*, 2014, **27**, 447-455.