PCCP

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

Cite this: *Phys. Chem. Chem. Phys.*, 2023, **25**, 11908

The relevance of short peptides for an understanding of unfolded and intrinsically disordered proteins

Reinhard Schweitzer-Stenner

Received 30th January 2023,
Accepted 18th April 2023

DOI: 10.1039/d3cp00483j

rsc.li/pccp

Over the last thirty years the unfolded state of proteins has attracted considerable interest owing to the discovery of intrinsically disordered proteins which perform a plethora of functions despite resembling unfolded proteins to a significant extent. Research on both, unfolded and disordered proteins has revealed that their conformational properties can deviate locally from random coil behavior. In this context results from work on short oligopeptides suggest that individual amino acid residues sample the sterically allowed fraction of the Ramachandran plot to a different extent. Alanine has been found to exhibit a peculiarity in that it has a very high propensity for adopting polyproline II like conformations. This Perspectives article reviews work on short peptides aimed at exploring the Ramachandran distributions of amino acid residues in different contexts with experimental and computational means. Based on the thus provided overview the article discussed to what extent short peptides can serve as tools for exploring unfolded and disordered proteins and as benchmarks for the development of a molecular dynamics force field.

1. Introduction

For a long period of time one of the central dogmas of protein biochemistry stipulated that to perform a biological function a polypeptide chain has to fold into a specific structure. For a monomer, this structure is the three-dimensional arrangement of so-called secondary structures (helices, β-strands and sheets, β and γ-turns) connected by unordered loop segments where a segment adopts none of the canonical secondary structures while individual residues adopt a specific backbone structure. This arrangement is generally termed tertiary structure. For some functions, monomers become the subunits of highly symmetric blocks of proteins which constitute the quaternary structure.[1] Human hemoglobin is a canonical example where functionality involves an interplay between changes of the tertiary and quaternary structure.[2,3]

While the relationship between protein structure and function has been the focus of biochemical and biophysical research for a long period of time, the respective unfolded state has attracted only scant attention. This attitude rooted in the belief that the high degree of disorder depicted by unfolded proteins is mostly independent of the amino acid residues composition and that the manifold of sampled conformations can be described properly by the random coil model imported from polymer physics (*vide infra*).[4–7] This view of structure as a requirement for function and the biological irrelevance of unfolded proteins was severely challenged by the discovery of intrinsically disordered proteins (IDPs) capable of performing multiple functions mostly in a cellular context.[8–11] Disordered means that each residue sample different backbone conformations similarly to what happens in unfolded proteins. In addition to completely disordered proteins many partially folded proteins with functionally relevant disordered regions (IDR) have been identified.[12,13] In this context so called short linear motifs (SLiMs) play an important role in molecular recognition processes.[14–16] Estimates of the fraction of biologically relevant IDPs vary. In eukaryotic cells up to 30% of eukaryotic proteins are at least partially disordered.[17] In many cases IDPs or disordered segments of otherwise folded proteins are involved in molecular recognition processes which involve disorder → order transitions.[12,13,18,19] In less frequent cases a reverse process occurs, namely an order → disorder transition.[20] IDPs are also involved in the self-assembly of proteins into oligomeric, protofibrillar and fibrillar structures that have been implicated in several neurological diseases. Alternatively, interactions between IDPs can cause a phase separation and the formation of membraneless organelles.[21–23]

One of the questions that must be answered for a thorough understanding of IDPs is to what extent their structural properties resemble that of unfolded proteins. Fig. 1 compares the occurrences of amino acid residues in IDRs (IDPs) and in

Department of Chemistry, Drexel University, Philadelphia, PA 19104, USA.
E-mail: rs344@schweitzer-stenner.com

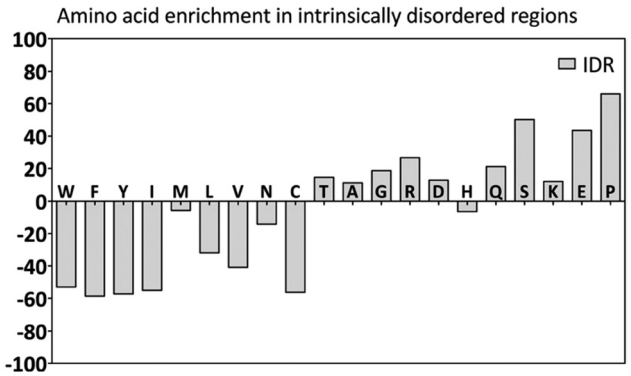

Fig. 1 Amino acid residue enrichment in intrinsically disordered regions of proteins. The enrichment is displayed as 100 − (%amino acid residue in IDRS × 100% × amino acid residue content). Taken from ref. 24.

globular proteins. The diagram reveals that all residue types with polar as well as positively (R, K) and negatively charged (D, E; at neutral pH) side chains occur in excess in IDPs and IDRs.[11,24,27] Obviously, this is not the case for unfolded states of foldable proteins. However, if as assumed by the canonical random coil model the conformational ensemble of unfolded proteins is sequence independent one is tempted to assume the same for IDPs and intrinsically disordered segments.

From the very beginning of protein biophysical and biochemical research very short peptides have served as model systems for exploring the conformational space that proteins can sample.[6,25,28–30] In this context the alanine dipeptides has played a major role for a long period of time.[30–35] Fig. 2 shows the Ramachandran plot for the backbone coordinates of the alanine residue in *N*-acetyl-alanyl-*N*-methylamide. This plot is just based on an exploration of steric and electrostatic interactions. It displays the sterically allowed and the favored regions for the backbone coordinates of the alanine residue. The depicted distribution should be compared with a Ramachandran plot representing adopted conformations of all types of residues (with the exception of proline and glycine) (Fig. 2). Obviously, the distributions depicted in Fig. 2 are very similar which leads to the conclusion that it is representative for all non-glycine and non-proline residues. Hence, one arrives at the conclusion that steric constraints and electrostatic effect are sufficient to describe the conformationally accessible space of polypeptide/protein residues in the unfolded state. In this context the alanine dipeptide can be used as a benchmark system, a sort of hydrogen atom for amino acid residues. Over time the modeling of alanine dipeptides has become more sophisticated in that more advanced force fields and solvent models were utilized. While all these studies agree that an aqueous solvent has a substantial influence on the Ramachandran distribution of alanine the resulting distributions were significantly different. On a qualitative level they can be divided into two types of conformational distributions. One type is very heterogeneous and suggest a broad nearly isoenergetic region in the upper hand quadrant of the Ramachandran plot (Fig. 2), which was entirely denoted β-strand. Conformations sampling this region are lower in energy than the ones in the right-handed and left-handed helical region. According to Tobias and Brooks this energy difference is substantially reduced due to peptide–solvent interactions.[36] The second type of studies yielded more structured and less inhomogeneous distributions with basins in the canonical β-strand region (φ-values below -100°) and in a region generally associated with the polyproline II (pPII) conformation (Fig. 3).[34] For some force fields and water models pPII becomes the most stable conformation. These comparatively recent results reflect an experimentally supported trend in the literature which suggest that the classical random sampling of the sterically allowed region in the Ramachandran plot is too simplistic. It is the goal of this

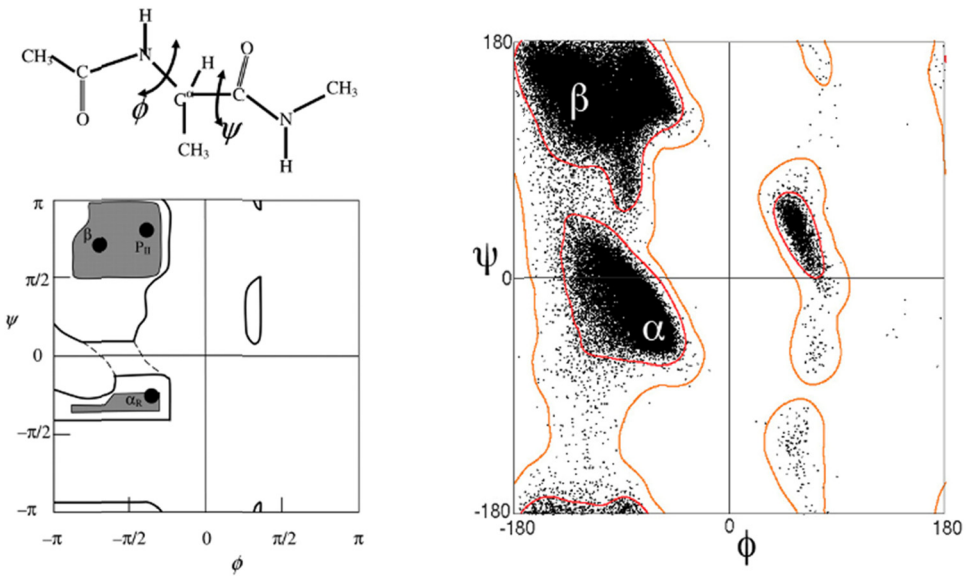

Fig. 2 Left: Schematic representation of the sterically allowed region of the Ramachandran plot of the depicted alanine dipeptide. Taken from Hermans,[25] open access. Right: Ramachandran plot of 10^5 residues in published protein structures produced by J. S. Richardson.[26]

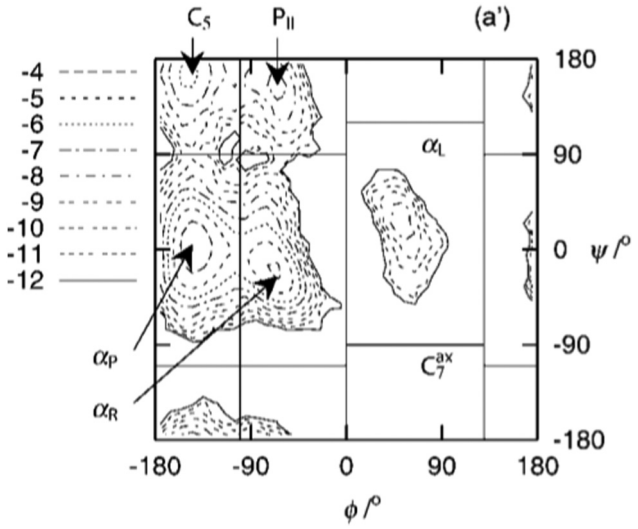

Fig. 3 Contour plot of the Ramachandran probability distribution of an alanine dipeptide in water obtained from MD simulations with an Amber ffParm 99 force field. The indicated contour values represent the logarithm of the normalized probability function. Individual basins are related to the indicated secondary structure conformations (C_5: extended β-strand like, P_{II}: polyproline II, α_R and α_L: right- and left handed α-helical, α_p: right-handed helical conformation close to the π-helix region, C_7^{ex}: region of the Ramachandran plot associated with (sterically forbidden)) inverse polyproline II and β-strand conformations. Reprinted with permission from ref. 34, 2008, American Chemical Society.

Perspectives article to outline the use of short peptides for the development of a more realistic picture of Ramachandran distributions of individual residues which differ much more from each other than expected for a long period of time. These peptides allow for an elucidation of conformational distributions of amino acid residues in the absence of non-local (mostly hydrophobic) interactions, which one can still expect to occur in unfolded proteins. The comparison of peptides of different length allows for the exploration of how nearest neighbors affect conformational distributions.

This article is structured as follows. Section 2 will be used to state the problem, namely the applicability of the random coil concept to unfolded and disordered peptides and proteins. In this context we will follow an earlier articulated concept that distinguishes between local and global aspects of the random coil theory.[37] Section 3 provides an overview of published experimental data that were interpreted as indicating that alanine has a high propensity for pPII conformations in oligopeptides. In Section 4, we describe the results that emerged from an extension of these structure analyses to non-alanine residues in blocked peptides including dipeptides. Section 5 provides a summary of investigations on unblocked tripeptides that utilized a broad data set of NMR and vibrational spectroscopy data (in blocked peptides the terminal carboxylic acid and ammonia groups of unblocked peptides are generally replaced by esters and methyl groups). The contribution of water to the stabilization of backbone conformations is briefly discussed in Section 6. Section 7 introduces the concept of nearest neighbor interactions to account for the observed context dependence of Ramachandran plots of amino acids even in short peptides. Section 6 discusses how results obtained with short peptides could be used in the future for force field development and for an understanding of local order in conformational entropy of unfolded and intrinsically disordered proteins. A Summary and outlook section finishes this article.

2. Stating the problem: is the state of unfolded proteins and IDPs a random coil

Even a superficial screening of the literature will inform the reader that the term 'random coil' is being used as a synonym for an unfolded state. If for instance, researchers observe UVCD spectra like those shown in Fig. 4 which can be described as a superposition of a negative Cotton band (below 200 nm) and a

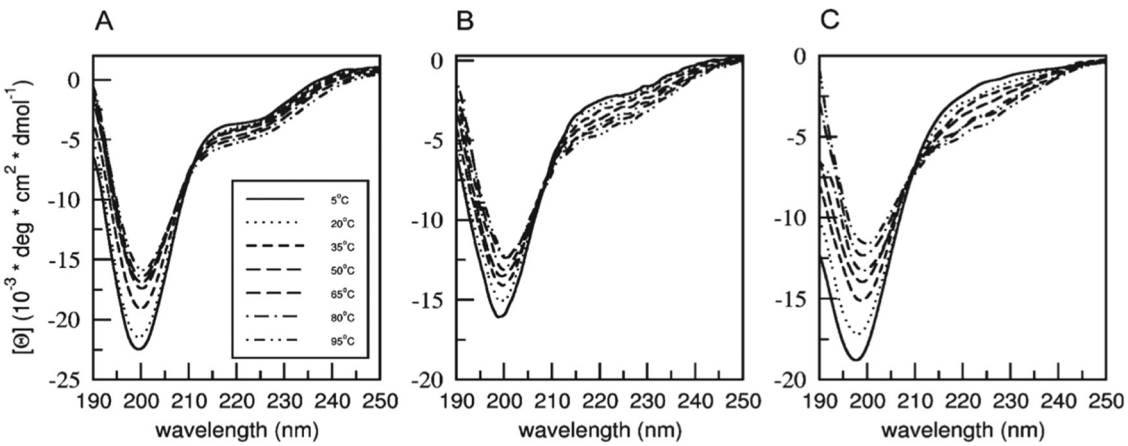

Fig. 4 UVCD spectra of three disordered protein segments measured at the indicated temperatures. Left: Activation domain of the thyroid hormone activator, (center) the cytosolic C-terminal distal tail of the human sodium–proton exchanger; (right) the S-phase delayed protein. Reprinted from ref. 38, open access.

shallow saddle point between 210 and 220 nm they interpret it as random coil indicator without paying much attention to different intensities in spectra of different proteins.[39] Other spectroscopic indicators (amide I wavenumber at 1640 cm^{-1} in IR-spectra,[40] chemical shifts of amide protons and ^{13}C that are close to the ones observed for short peptides) are generally interpreted in a similar manner.[41] Strictly speaking, however, the term random coil solely applies to long polymers formed with rigid building clocks (peptide groups) and freely rotatable linkers.[7,42] Its length dependence can be described by a power law for the radius of gyration and the mean radius of hydration, *i.e.* $\langle R_h \rangle \backsim N^{\nu}$. If different proteins behaved as an ideal random coil, the exponent would assume the random walk value of 0.5. A more realistic self-avoiding random walk model which takes the excluded volume into account predicts an exponent of 0.59.[42]

Deviations from these scaling laws are possible and have been frequently observed. The self-avoiding random walk (excluded volume) model seems to be suitable for proteins denatured in urea.[43] However, the behavior of unfolded proteins and IDPs depends very much on the respective solvent–protein and intramolecular interactions. Only if both interactions are perfectly balanced and exactly cancel out does the protein exhibit ideal random coil behavior (θ-point). If protein–protein interactions exceed protein–solvent interactions (poor solvent), the unfolded or disordered protein adopts a more compact structure with much less conformational flexibility. Consequently, the exponent ν becomes significantly smaller than 0.5.[42] On the contrary, in a good solvent protein–solvent interactions would be predominant and the exponent would exceed 0.5. For foldable proteins, water is a poor solvent at room temperature. The addition of large amounts of urea and guanidine chloride (GmdCl) generally denatures folded proteins. A common scaling factor of 0.59 seems to indicate a self-avoiding random coil in a good solvent, though this notion has recently been questioned by Holehouse *et al.*, who showed that denaturation in urea and GmdCl involves a combination of side chain and preferential binding effects.[44] Regarding IDPs which normally contain a higher fraction of ionizable residues than foldable proteins[11,45] it is important to note that the value of the exponent ν increases with increasing net charge and can even exceed 0.6.[46,47]

The above discussion reflects the global aspect of the random coil theory, namely the size dependence of the polymer as a function of internal and external parameters. Locally, the random coil model assumes that rigid building blocks are connected by freely rotatable links. In the case of polypeptides the rotatable links are the N–C_α and C_α–C′ bonds of an amino acid residue.[37,42] As we know from the above introduced Ramachandran plots the rotational motions around these bonds (associated with the dihedral angles φ and ψ) are restricted mostly for steric reasons. However, the sterically allowed conformational space as shown in Fig. 1 is large enough to justify that locally unfolded peptides behave like random coils, irrespective of side chain composition and the choice of the solvent.

While the evidence in favor of the random coil concept seems to be overwhelming, experimental and bioinformatical evidence gathered over the last twenty-five years cast some doubts on its full applicability to unfolded proteins and IDPs. First, an ideal random coil state would not depict any residual structure. However, some very thorough NMR studies on denatured proteins and IDPS such as α-synuclein and tau provided compelling evidence for the occurrence of local (transient) helical and sheet structures most likely facilitated by non-local intra-protein interactions.[48–51] Second, starting with some (at that time) provocative work on oligo-alanine peptides, evidence has been gathered for conformational preferences of individual amino acid residues.[37,52,53] These results suggest that the conformational ensemble sampled by an unfolded protein or IDP depends on the amino acid residue composition. The dependence of power law exponent on the net charge already points in this direction.[46,47,54] However, current theories treat charges as increasing the excluded volume of a polypeptide chain rather than worrying about their influence on individual Ramachandran distributions.[46,54] Overall, the above mentioned results imply that the conformational entropy of IDPs might be less than what one would expect for an ideal random coil which would be of importance for any modeling of protein dynamics in water. In what follows in the next section an overview is provided of how short peptides have been used to explore intrinsic conformational propensities. Results of these studies are compared with related studies of coil libraries.

3. Conformational preferences of amino acid residues I: oligo-alanine peptides

Two papers that both appeared in 2002 triggered a discussion and various types of investigations of intrinsic properties of amino acid residues. The first one was published by Kallenbach and colleagues.[55] They investigated an oligopeptide with the sequence $AcX_2A_7O_2$-NH_2 (X: aminobutyric acid, O: ornithine) termed XAO. The authors measured the UVCD spectra of the peptide as a function of temperature. The shape of the spectrum at room temperature (Fig. 5) resembles the one in Fig. 4, so a conventional view would interpret it as indicating a random coil. However, the very pronounced temperature dependence and the apparent existence of an isodichroic point, which is diagnostic of a two-state transition, argue against such a view, in particular because the difference spectrum in Fig. 5 looks very much like that of a β-strand or β-sheet conformation. In addition, the authors extracted $^3J(H^N H^{C\alpha})$ for all seven alanine residues from the ^1H NMR spectrum of the peptide. This coupling constant can be obtained from the splitting of amide proton signals. Its dependence on the dihedral angle φ can be described by a Karplus equation the most general form reads as:

$$J(\eta) = A\cdot\cos^2(\eta + \theta_1) + B\cdot\cos(\eta + \theta_2) + C \quad (1)$$

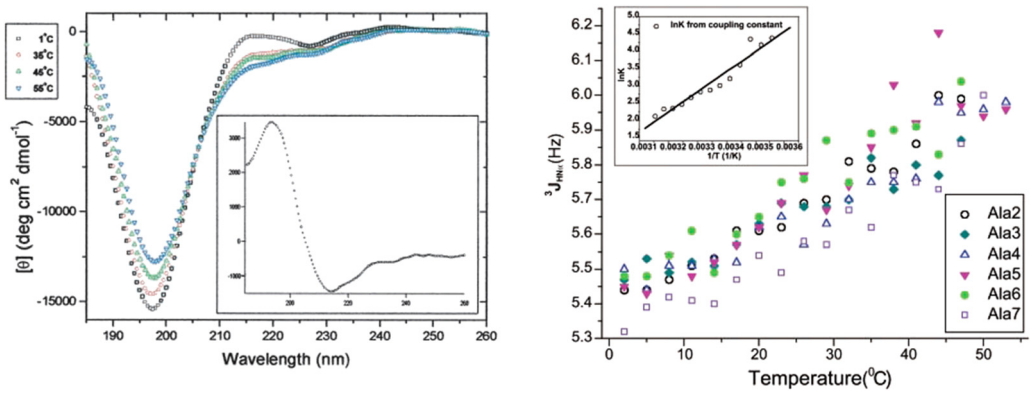

Fig. 5 Left: UVCD spectrum of the heptapeptide XAO measured at the indicated temperatures. The inset shows the difference spectrum calculated by subtracting the spectrum measured at 1 °C from the one measured at 55 °C. Note that the *y*-scale which is already blurred in the original figure ranges from −100 to 300 deg cm^2 dmol^{-1}. Right: $^3J(H^NH^{C\alpha})$ coupling constants of the indicated alanine residues plotted as a function of temperature. The inset exhibits a plot depicting the result of a two-state van't Hoff analysis described in ref. 55 from where the figure was taken (open access).

where $\eta = \varphi, \psi$, depending on which type of coupling constant is analyzed. The amplitudes A, B, C and the phases θ_i are empirical parameters that researchers have obtained from fits to *J*-coupling constants observed for proteins for which high quality crystal structures or NMR-based structures are available.[56,57] A discussion of these parameters and their uncertainties is given in Section 8. The experimental values Shi *et al.* obtained for the seven alanine residues cluster all around 5.5 Hz at room temperature (Fig. 5). Such a value suggests that their conformational distributions reflected by the measured average $^3J(H^NH^{C\alpha})$ value at room temperature are dominated by the sampling of right-handed helical and/or in the polyproline II (pPII) region of the upper left quadrant (*cf.* Fig. 1). Since the UVCD data ruled out the former, the authors opted for the latter and reported that alanine predominantly samples the pPII region of the Ramachandran plot. The notion of a pPII dominance was further supported by nuclear Overhauser effect (NOE) measurements, which ruled out a major sampling of right-handed helical conformations. All $^3J(H^NH^{C\alpha})$ constants increase with temperature (Fig. 4), which is consistent with the notion of conformational redistribution from pPII to β-strand indicated by the CD spectra. The inset in Fig. 5 (right) exhibits a van't Hoff plot of the Gibbs energy difference between pPII and β-strand extracted from the depicted $^3J(H^NH^{C\alpha})$ data by employing a simple two-state model.

Generally, pPII is a structure associated with poly-L-proline with all its peptide group in the *trans*-conformation.[58] In its crystalline state the corresponding φ and ψ-values are -75° and 150°. For proline residues, it is highly preferred for steric reasons. When the XAO data were published there did not seem to be any obvious reason why alanine should prefer the same conformation.

The conclusions drawn from the XOA study imply that its UVCD spectrum is diagnostic of a pPII conformation and not of a random coil supporting distribution. This notion agrees with the fact that poly-L-proline shows a very similar UVCD spectrum, just with its extrema at slightly different positions. This similarity between the polyproline II CD spectrum and the spectra of ionized poly-L-lysine and poly-L-glutamic acid had been noticed at an early stage by Tiffany and Krimm, who arrived at the conclusion that the unfolded state of these peptides contains a predominant fraction of pPII.[59] After a controversial debate[60,61] the scientific community decided to ignore such an inconvenient truth.

The second work that must be mentioned in this context is the femtosecond two-dimensional IR study of Woutersen and Hamm on trialanine in acidic aqueous solution.[62] Femtosecond pump–probe experiments allowed them to determine the strength of the excitonic coupling between the amide I′ modes of the two peptide groups which in D_2O and the angle between the transition dipole moment of these amide modes. Thus, the authors identified a representative structure at φ and ψ values of -60° and 140° which puts it right into the pPII region, thus confirming the results of Shi *et al.*[55]

The results of Shi *et al.* provoked a very controversial debate which mostly focused on the interpretation of their experimental results.[63,64] Interestingly, the critics of this study mostly overlooked the confirming results of Woutersen and Hamm. I am referring the interested reader to earlier reviews that provides more details of the debate.[37,53,65] Here, I confine myself on spectroscopic studies that resolved the debate very much in favor of Shi *et al.*

Graf *et al.* used a set of seven NMR scalar coupling constants to determine the conformational distribution of several unblocked oligo-alanine peptides, including trialanine (A_3).[66] In addition to $^3J(H^NH^{C\alpha})$ the authors utilized the φ-dependent $^3J(H^NC')$, $^3J(H^{C\alpha}C')$, $^3J(C'C')$, $^3J(H^NH^{C\beta})$, the ψ-dependent $^1J(N,C_\alpha)$ and $^2J(N,C_\alpha)$ and the φ and ψ dependent $^3J(H^NC_\alpha)$ coupling constants. Fig. 6 depicts several Karplus curves for the φ-dependent ^3J coupling constants. The different Karplus curves result from empirical fits to different data sets and from DFT-calculations for alanine residues. Graf *et al.* employed the parameters reported by Hu and Bax.[57] Since the Karplus curves for the utilized ^3J-coupling constant are very different, their combined use enables a reliable assessment of conformational distributions along the φ-coordinate axis. In addition

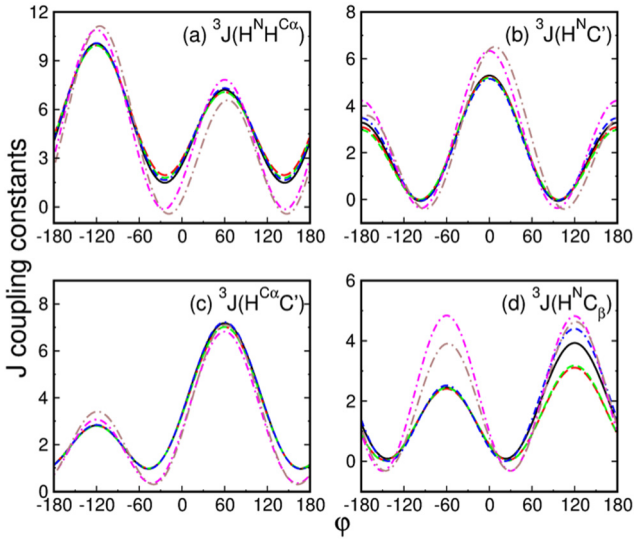

Fig. 6 Karplus curves of the indicated φ-dependent J-coupling constants calculated with different Karplus parameters. Solid black and red dashed: Hu and Bax,[57] dashed green: Wang and Bax,[56] derived from crystal structures, blue dashed-dot, Wang and Bax,[56] from X-ray and NMR data, dashed-dot green and dark green: Case *et al.*, DFT1 and DFT2,[67] respectively. Reprinted with permission from ref. 68, 2020, American Chemical Society.

the ψ-dependence of $^1J(N,C_\alpha)$ facilitates the differentiation between extended (pPII and β-strand) and right handed helical conformations. In order to analyze their data, the authors first used molecular dynamics simulation with a GROMOS96 force field and a SPC water model to obtain basins in the pPII, β-strand and right-handed helical region of the Ramachandran plot. In a second step the authors used the mole fractions associated with these three basins as free parameters in a fit to the experimental coupling constants. For the central residue of A_3, the obtained mole fractions were $\chi_{pPII} = 0.92$, $\chi_\beta = 0.08$ and $\chi_\alpha = 0$. For longer oligo-alanine peptides the results suggest a slight stabilization of β-strand. For A_7, χ_{pPII} varies between 0.83 and 0.86. The population of right-handed helical states was found to be negligible for all non-terminal alanine residues. Thus, the work of Graf *et al.*,[66] by using a much broader data set than the preceding works, fully confirmed the results of Shi *et al.*[55] They added considerable value to the debate by their quantitative determination of conformational propensities.

4. Conformational preferences of amino acid residues II: comparison of residues in blocked glycine-based host–guest peptides

The results of the studies on alanine-based peptides led researchers to wonder whether similar deviations form random coil supporting behavior exist for other amino acid residues. Three different experimental investigations on blocked short (and ultrashort) peptides have addressed this issue. Kallenbach and associates investigated a complete guest series of the

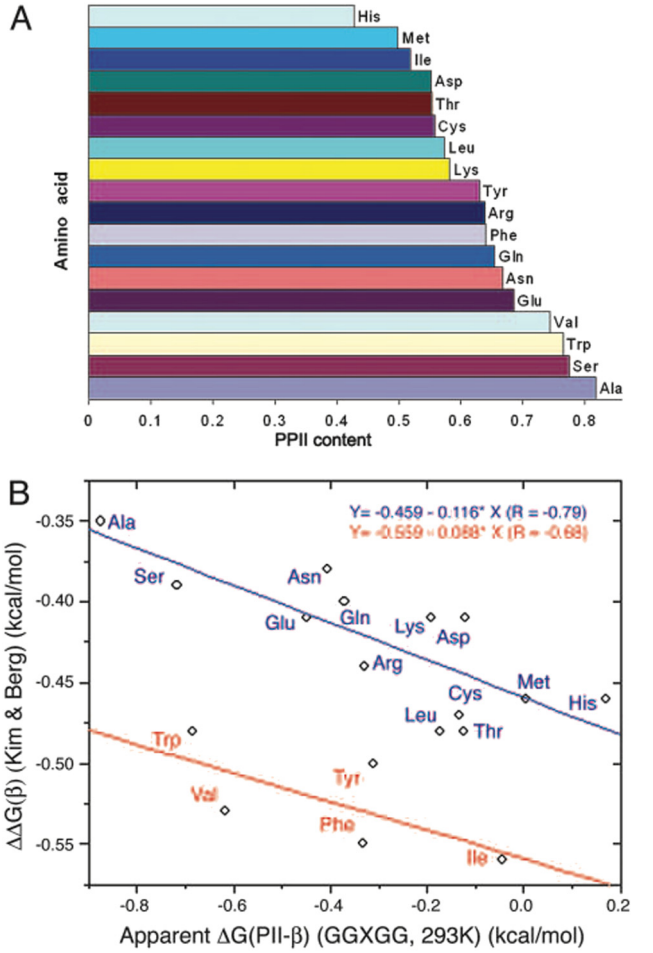

Fig. 7 (A) Bar diagram depicting the pPII-fraction of the guest residue in Ac-G_2xG_2-NH_2 a derived from the respective $^3J(H^NH^{C\alpha})$ coupling constant as explained in the text. (B) Correlation between the Gibbs energy difference between the pPII and β-strand structures of the indicated guest residues and β-sheet propensities reported by Kim and Berg.[181] The figure was taken from ref. 69 (open access).

oligoglycine $AcG_2xG_2NH_2$ (abbreviated as G_2xG_2 in the following), where x presents one of 19 natural amino acid residues (x = G was not investigated).[69] These authors used again the experimental protocol of their XAO studies. With the exception of the peptides with x = H, Y, W and F all UVCD spectra qualitatively resembled the one observed for XAO. The deviations for aromatic residues can be expected owing to the electronic interactions between backbone and side chain transitions.[70,71] Since the CD spectra exhibit isodichroic points, the authors assumed that the different $^3J(H^NH^{C\alpha})$ coupling constants observed for different x, which span from 5.7 to 7.8 Hz, solely reflect the presence of two states. For their analysis, they obtained representative $^3J(H^NH^{C\alpha})$ coupling constants for the pPII and β-strand conformations of individual residues from coil library distributions reported by Aveblj and Baldwin.[72] Thus, they obtained the pPII propensity diagram in Fig. 7 which suggests that only the Ramachandran space of histidine is not dominated by pPII. A total of 4 residues (A, S, V and W) have propensities for pPII above 0.7 which puts the

Gibbs energy differences between pPII and β close to RT at room temperature. 11 residues exhibit pPII propensities above 0.6. Alanine is on top of the list with 0.83. Hence, these results suggest that the Ramachandran plots of individual amino acid residues are mostly dominated by pPII and that helical contributions are negligible.

A different approach has been undertaken by Grdadolnik *et al.* who investigated 19 different amino acid dipeptides.[73] In addition to the respective $^3J(H^NH^{C\alpha})$ coupling constants they utilized the band profile of amide III in the Raman and IR spectra of the investigated peptides. They assigned three subbands underlying the amide III profile to pPII, α-helical and β-strand. Fig. 8 illustrates the pPII, β- and right-handed helical fractions reported by Gradadolnik *et al.* The pPII population spreads from 0.6 for alanine (significantly lower than the value Shi *et al.*[55] reported for G_2AG_2) to 0.38 for histidine. The helical fractions are generally weak (below 0.1). The β-strand fraction is considerable for H, protonated D, N, T, C, I and V (all above 0.5). Thus, the values reported by Grdadolnik at al. match much more expectations that the ones of Shi *et al.*, since they are more in line with established propensities for β-sheets.[74]

Despite the quantitative differences between the propensity values that emerged from the above studies their results both indicate that the conventional wisdom of a residue independent Ramachandran distribution of amino acid residues does not withstand experimental scrutiny. Both studies agree in suggesting that all residues sample predominantly pPII/β-strand in the upper quadrant of the Ramachandran plot, which is at variance with rather high population of right-handed helical structures deduced from models that solely consider steric hindrance and electrostatic effects.[4,6,28,29] This implies that locally residues are less random than assumed for the random coil model.

While the works discussed in this section deserve credit for shedding light on intrinsic structural properties of amino acid residues in aqueous solution, they leave several issues unresolved. First, all the respective NMR analyses of $^3J(H^NH^{C\alpha})$ constants were based on the unproven assumption that the centers of the basins for pPII and β-strand in the Ramachandran space of coil libraries are representative for short peptides in solution. As shown below this is not the case (Section 8). Second, the conventional Ramachandran distributions of amino acid residues were replaced by just two (Shi *et al.*[55]) or three points in the configuration space (Grdadolnik *et al.*[73]) spanned by the backbone dihedrals. This would not be realistic even for folded proteins. Third, it seems to be unlikely that residues solely sample the upper left quadrant of the Ramachandran plot. Fourth, relying predominantly on a single *J*-coupling constant is problematic since the respective Karplus curve suggests several solutions for the same coupling constant value (*vide infra*). Fifth, in spite of its convincing results the spectral analysis of the amide III profile carried out by Gradadolnik *et al.* ignores the multiplet structure of this band which arises from vibrational mixing with CH bending modes of the backbone and particularly aliphatic side chains.[75–77]

5. Conformational preferences of amino acid residues III: comparison of residues in unblocked tripeptides

In addition to their analysis of oligo-alanine peptides Graf *et al.*[66] also investigated the conformational distribution of the central residues in the tripeptide V_3 in order to determine to what extent hydrophobicity and steric demand of a side chains matter regarding the population of different basins in the Ramachandran plot. A comparison of mole fractions of alanine (in A_3) and valine is shown in Fig. 9. Apparently, the conformational distribution of valine is quite different form the one of alanine. The authors obtained fractions of 0.29, 0.52 and 0.19 for pPII, for β-strand and right-handed helical conformations of valine, respectively. These values are also clearly distinct for the one that Shi *et al.* reported for G_2VG_2. The discrepancy is less pronounced for the distribution Grdadolnik *et al.*[73] reported for the IR/Raman based valine dipeptide values (0.47, 0.51, and 0.02).

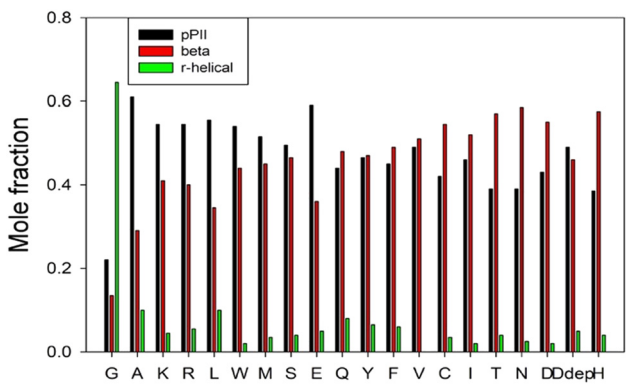

Fig. 8 Bar diagram depicting the mole fractions of pPII, β-strand and right-handed helical conformations of the indicated dipeptides. The data were taken from ref. 73. The values represent averages of the mole fractions obtained with IR and Raman spectra of the investigated peptides.

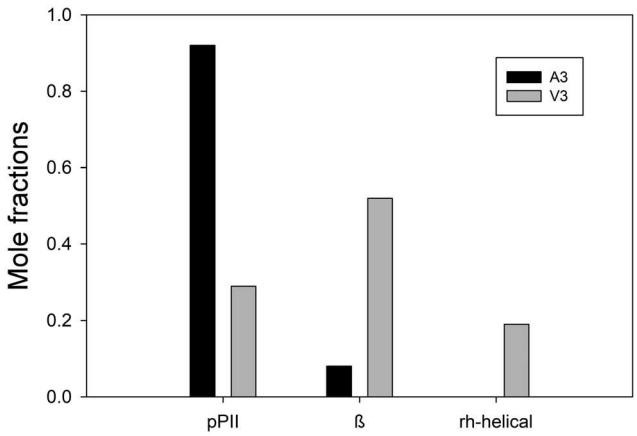

Fig. 9 Conformational propensities for the indicated conformations sampled by the central residue of cationic A_3 and V_3. The data were taken from Graf *et al.*[66]

The above work of Graf *et al.* triggered a series of investigation of tripeptides that combined their use of NMR coupling constants with the analysis of amide I′ profiles in IR, polarized Raman and vibrational circular dichroism (VCD) spectra. Amide I′ (the prime symbol indicates that the amide group is deuterated) is predominantly a CO stretching mode.[78] In a polypeptide chain amide I modes are coupled *via* orientation dependent electrostatic interactions. They cause a delocalization of the excited vibrational states and thus a change of amide I′ band profiles and positions.[79,80] In order to simultaneously analyze amide I profiles and *J*-coupling constants Schweitzer-Stenner constructed Ramachandran plots as a superposition of two-dimensional Gaussian functions positioned at basins of the Ramachandran plot[81] (termed Gaussian model in the following). Positions, halfwidths and statistical weights were used as adjustable parameters. The author combined the amide I′ profiles of cationic A_3 and V_3 with the *J*-coupling constants of Graf *et al.* to obtain the Ramachandran plots in Fig. 10. Bar diagrams in Fig. 10 (right) compare mole fractions obtained from this analysis with the ones of Graf *et al.*[66] With regard to A_3, the results were very similar. The pPII fraction reported by Schweitzer-Stenner is slightly lower (0.84). In addition to pPII and β-strand (0.08) he identified small populations of right-handed helical and inverse γ turn structures (0.04 each). Despite these minor differences this analysis confirmed the notion that alanine has an unexpectedly high pPII propensity. For valine, the results of Schweitzer-Stenner suggest a higher β-strand propensity (0.68) than Graf *et al.* Hence, his results further widened the gap between the conformational distributions of alanine and valine.

He *et al.* argued that the use of unblocked peptides might be problematic because electrostatic end effects could influence to the conformational distributions at least for very short peptides such as A_3 and A_4,[83] thus reiterating earlier reported skepticism.[84] The authors cited the fact that four guest residues in GxG, $AcGxGNH_2$, and $AcGGxGGNH_2$, and the respective dipeptides show slightly different $^3J(HNH\alpha)$ coupling constants at different pH as an argument for the influence of terminal groups.[83]

This issue was addressed by Toal *et al.*, who compared the structural distributions of the three protonation states of A_3 and combined their analysis with MD simulations.[85] For the conformational analysis of A_3, they employed the above-described combination of *J*-coupling constants and amide I′ band profiles. In addition to A_3 the authors analyzed data for the alanine dipeptide. They could draw the following conclusions from their data: first, the influence of the terminal charges on the conformational distribution of the central residue of A_3 is negligible. Second, the conformational distribution of the alanine dipeptide resembles that of alanine residue in GAG, for which the pPII fraction is slightly lower than observed for the central alanine residue in AAA. Third, the results of the MD simulations strongly suggested that pPII of alanine is stabilized by backbone and side chain hydration, in line with results of earlier computational studies.[86–89] The relationship between conformational propensities and hydration is discussed in more detail in Section 6.

To obtain a more complete picture of how conformational propensities of amino acid residues depend on the characteristics of the side chains Schweitzer-Stenner, Schwalbe and associates investigated a representative series of cationic GxG peptides by combining NMR and vibrational spectroscopic measurements.[90–93] The data were analyzed with the above introduced Gaussian model. Results of the studies have been the subject of earlier reviews.[37,53] Fig. 11 compares the obtained mole fractions for pPII and β-strand with corresponding values of G_2xG_2 and respective dipeptides. Most of the GxG data were obtained with multiple coupling constants. Contrary to Shi *et al.*,[69] the values depicted in Fig. 11 suggest that besides alanine, only a very limited number of amino acid residues exhibit pPII propensities above 0.5, namely M, L, E, C and R. Results of the above work on dipeptides suggest that the pPII propensities of K, R, L, M and E exceed this value, which demonstrates substantial overlap between the two studies. Two particularly remarkable results of the GxG work should be emphasized here. First, the extremely low pPII propensity of protonated GDG (it is slightly higher in the ionized state) and second, the population of turn-supporting conformations for GxGs with side chains capable of either donating or accepting hydrogen bonds.[90,93] The total fraction of the pPII-β-strand population varies between 0.7 and 0.9.

In order to assess the significance of the displayed numbers Fig. 12 plots the Gibbs energy difference between pPII and β-strand as a function of the pPII fraction for different total

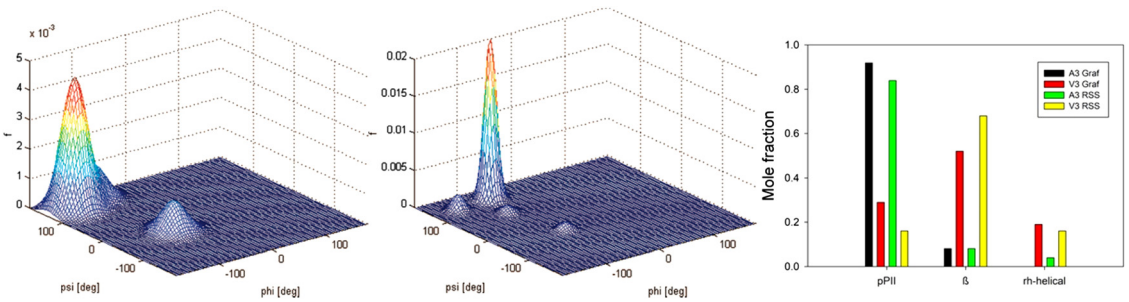

Fig. 10 Three-dimensional Ramachandran plots of the central residues of cationic A_3 (left) and V_3 (middle) obtained from a global analysis of *J*-coupling constants reported by Graf *et al.* and amide I profiles of Eker *et al.*[82] The plots were reprinted with permission from ref. 81, 2009, American Chemical Society. Right: Bar diagram comparing the mole fractions of the central residues of A_3 and V_3 reported by Graf *et al.*[66] and Schweitzer-Stenner.[81]

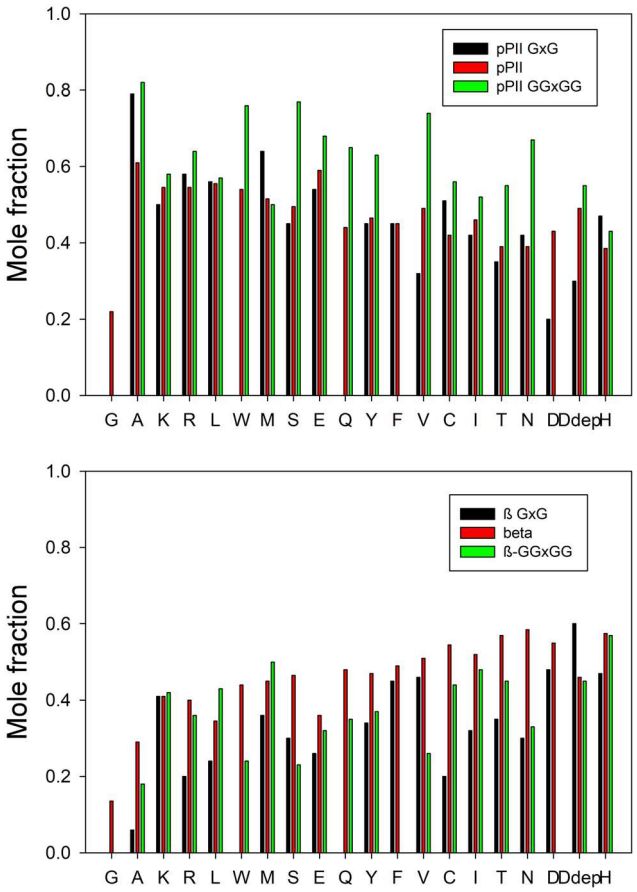

Fig. 11 Bar diagrams comparing the mole fractions of pPII and β-strand obtained for GxG (black), blocked dipeptides (red) and the host residue of G_2xG_2 (green). Data were taken from ref. 69, 73 and 90–93.

fractions of residues sampling the upper left quadrant of the Ramachandran plot. In all four curves plotted in Fig. 12 the Gibbs energy decreases from *ca.* 5.8 and to −5.8 kJ mol^{-1} with an increase of χ_{pPII} from 0.1 to 0.9 (alanine) if only pPII and β-strand are sampled. For protonated D, the fraction of turn-supporting structures is comparatively high (0.23).[93] Thus, changing the pPII fraction from 0.7 to 0.2 (obtained value for GDG), moves the Gibbs energy from *ca.* 2.2 to −4.3 kJ mol^{-1}. These changes of the Gibbs energy are significant in that they exceed the thermal energy.

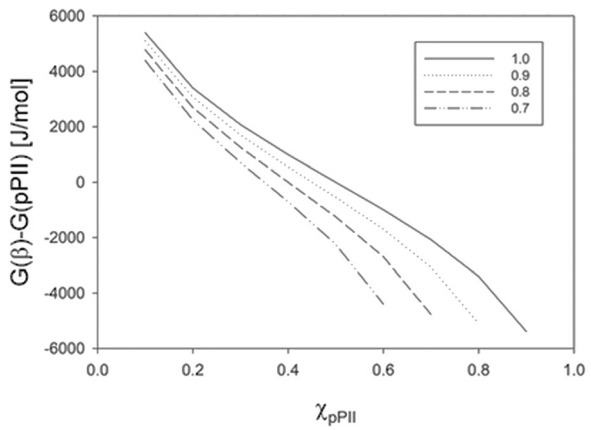

Fig. 12 Gibbs energy difference between pPII and β-strand as a function of the pPII fraction calculated for different fractions occupying the region above ψ = 100° in the right-hand half of the Ramachandran plot.

6. Conformational preferences of amino acid in coils libraries

Over the last twenty years coil libraries have emerged as an alternative source of data from which the conformational propensities of amino acid residues could potentially be determined. Coil libraries are constructed from dihedral angles of residues incorporated in unordered protein segments (*e.g.* loop regions).[94–99] The hypothesis behind this strategy is that any long-range interactions can be averaged out by using a large data set of such residues. To ensure a sufficient statistical quality, Ramachandran plots of individual amino acid residues are generally obtained by adding the data points for all nearest neighbors in the data set, thus ignoring the possible influence of nearest neighbors. To my best knowledge, only the publicly available coil library set of Sosnick and coworkers provides the means to obtain Ramachandran for GxG segments which could be directly compared with the above discussed experimental data.[100] The number of data points in the corresponding plots is rather limited. Noteworthy differences and similarities between x-distributions in GxG peptides and in the Sosnick library are discussed in the literature.[92] Here, I just mention the high pPII propensity of alanine and the extraordinary propensity of aspartic acid residues for turn-supporting structures (type I/II′ (i + 2) β-turn) which are both on display in the GxG and coil library-based Ramachandran plots.[101] As the experimental data obtained with short peptides coil library distributions reveal that Ramachandran distributions of amino acid residues can be quite distinct from each other. However, in most cases coil library distributions indicate a more pronounced sampling of right-handed helical structures than the experiment-based Ramachandran plots of corresponding GxG peptides.[92]

7. Relevance of conformational propensities of amino acid residues

How do the results of peptide studies discussed in this section affect our understanding of unfolded and disordered proteins? If one assumes the absence of nearest neighbor and non-local interactions, one arrives at the conclusion that their conformational entropy should be significantly lower than in the case of a random sampling of the sterically available conformational space.[97,102] Moreover, Gibbs energy differences between different peptide/protein conformations would be more pronounced than in the random coil case. Scheraga and coworkers, who recognized at an early stage that sterically allowed backbone conformations differ energetically, suggested to replace random coil with the term statistical coil.[103]

Two issues deserve to be addressed at this point. First, it should be emphasized that the thus far presented results on conformational propensities of amino acid residues should not be construed as indicating the occurrence of ordered secondary structures in unfolded and disordered states. Such a notion is not supported by the reported propensity values. Even if the pPII propensity of alanine is 0.9 as reported by Graf *et al.*,[66] the probability for the hepta-alanine segment of XAO to adopt a pPII helix would be just 0.39. As shown by Toal *et al.*,[85] conversions between pPII and β-strand conformation occurs on a picosecond time scale, which significantly curtails the lifetime of such a pPII segment. Hence, the use of the term 'pPII helices'[89,104–106] should be avoided. Some articles claimed that pPII helices formed by poly-alanines melt at higher temperature in the same way regular helices do.[107] While pPII helices are indeed formed if a polypeptide contains a lot of proline residues, there is no evidence that such a secondary structure can be adopted by unfolded/disordered proteins in the absence of any stabilizing non-local interactions. An example for the latter is the snow flea antifreeze protein, where 46% of the residues are glycines.[108] It does not have a classical hydrophobic core, yet it is fully folded with the structure comprising six pPII helices. The occurrence of pPII helical segments has been proposed for the N-terminal $A\beta_{1-9}$ based on $^3J(H^NH^{C\alpha})$ constants and the respective UVCD spectrum.[109] The temperature dependence of the latter was interpreted as suggesting a melting of the pPII helix into a random coil at high temperatures. However, as shown by Schweitzer-Stenner and Toal, the reported NMR and CD data can well be understood with a statistical coil model that considers nearest neighbor interactions (*vide infra*).[110]

The second issue is directly related to the proposed pPII helices. The claim of its existence in unfolded/disordered proteins led to the so-called reconciliation problem.[111–113] The argument reads as follows. If unfolded proteins are really composed of pPII-helical segments, wouldn't that imply a conflict with the experimentally verified random coil behavior of non-compact fully denatured proteins that generally obey a scaling law with an exponent of 0.59–0.6? Interestingly, however, Fitzkee and Rose demonstrated that the global behavior reflected by this exponent does not rule out an even heavily exaggerating model that describes an unfolded state as an ensemble of rods connected with flexible linkers.[113] Their result is important irrespective of the discussion about pPII helices in that it demonstrates the necessity to distinguish between local and global aspects of the random coil concept, as suggested earlier,[42] but it should be kept in mind that the proposed reconciliation problem does not exist.

8. Conformational propensities and hydration

While the work of Shi *et al.*[55] and the concomitant corroboration by Woutersen and Hamm[62] ignited the discussion about conformational propensities of amino acid residues in general and the pPII propensity of alanine in particular,[63,114] the latter had already been up in a less noticed paper by Han *et al.*[115] These authors used DFT calculation on the alanine dipeptide *N*-acetyl-L-alanine-*N*-methylamide to calculate vibrational spectra (Raman, VCD, Raman optical activity) for different conformations of the peptide. They were obtained by geometry optimization in implicit and explicit water. For the latter case, they considered four water molecules hydrogen bonded to the functional peptide groups (CO and NH). The authors found the stabilization of a novel structure in the presence of four water molecules, for which they obtained dihedral angles of $\varphi = -93.55^\circ$ and $\psi = 127.62^\circ$. Though somewhat different from the canonical pPII structure, it was close enough to earn this designation. In the absence of explicit water the γ-turn like structure C_7^{eq} emerged as the most stable conformation. In other words: this work already suggested that in water pPII might indeed be the most stable conformation that alanine can adopt in aqueous solution.

As mentioned above water had not been originally considered for the construction of Ramachandran plots.[4,29] Later MD simulations filled that gap but results seemed to be even more random coil like than the original distributions.[36] However, after the XAO results were reported a lot of computational work focused on the role of water. In a remarkable study Garcia performed MD simulations with a modified AMBER force field to show that the high preference of alanine for pPII in unblocked oligo-alanine peptides results from a favorable packing of water molecules around the peptide backbone.[87,116] A more specific picture arose from the MD studies of Mezei *et al.* who investigated conformational preferences of a 12-residue poly-L-alanine peptide with CHARMM 22 and TIP3P water.[117] They found that pPII is favored over three other conformations (antiparallel and parallel β-strand, right-handed helical) by backbone-water hydrogen bonding. In the β-strand conformation hydration water adopts entropically unfavorable bridge structure reminiscent of cages around hydrophobic groups. While compelling, this view is at variance with multiple thermodynamic studies on XAO oligo-alanines, G_2xG_2 and GxG that all clearly suggest that β-strand is entropically favored over pPII, while the latter is enthalpically favored.[66,69,118,119] A different view was presented by Avbelj and Baldwin based on electrostatic calculations.[120] They demonstrated the shielding role of hydration water which diminishes electrostatic interactions between peptide units which in the absence of water would prefer a more extended β-strand population.

DFT-based calculations for unblocked tripeptides in water strongly supported the view that water–peptide interactions stabilize pPII. In this context the work of Lanza and Chiacchio is particularly remarkable.[121–123] The authors investigated the role of hydration water regarding the stabilization of backbone conformations of cationic trialanine. They added a total of 37 water molecules to the peptide's hydration shell. Fig. 13 shows some of their peptide–water complexes. In addition to the central residue, they also considered the conformation of the C-terminal residue. They found that pPII–pPII dimers become

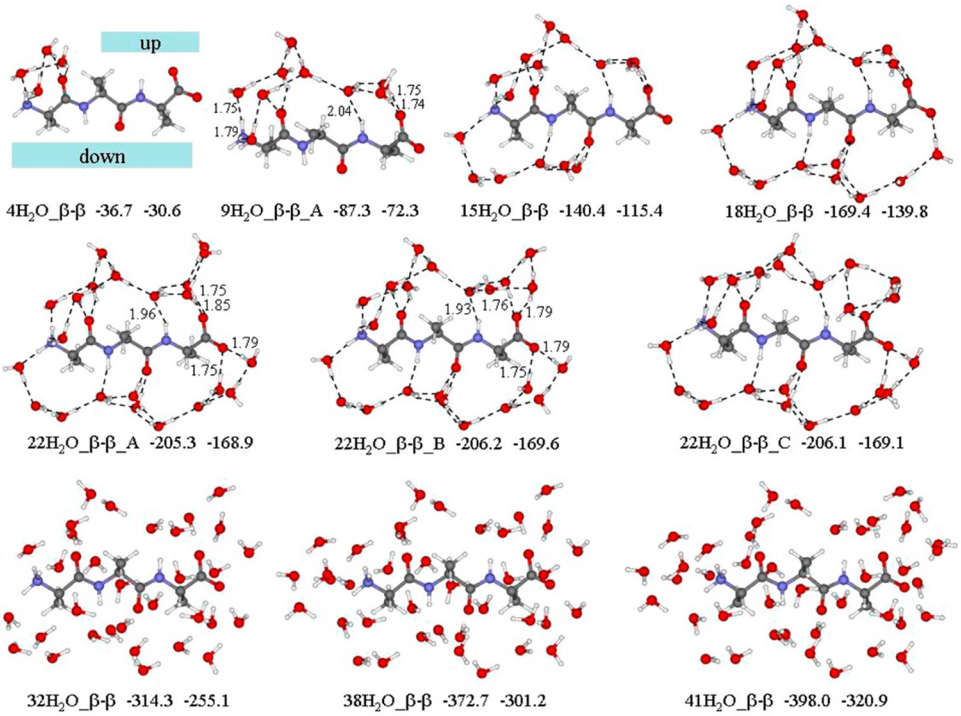

Fig. 13 Optimized molecular structures of A_3 imbedded in complexes of water molecules with the indicated numbers of water molecules. All peptides shown adopt a β-strand conformation. The numbers represent the corresponding internal energies in kcal mol^{-1} calculated with a 6-31+G* and a aug-cc-pVTZ level of theory. Reprinted with permission from ref. 121, 2016, American Chemical Society.

increasingly stabilized with increasing number of water molecules, which reflects a more efficient intermolecular hydrogen bonding. This notion was corroborated by more recent DFT calculations on four GxG peptides (x = A, V, L, I) where explicit hydration was only modeled with 10 water molecules.[124] This study was more limited in its goals in that it focused on pPII and β-strand conformations. It had been triggered by the surprising finding that the enthalpic and entropic difference between these two conformations is particularly large for the aliphatic residues V and I (~40 and 60 kJ mol^{-1}, respectively).[119] The respective Gibbs energy difference was found to be small (<1 kJ mol^{-1}) due to enthalpy–entropy compensation. The results of the DFT study reproduced the thermodynamic results at least on a qualitative level. They clearly revealed a stabilization of pPII *via* peptide–water interactions. Regarding both, enthalpy and entropy, vibrational mixing between peptide and water modes in the region below 700 cm^{-1} was found to be of particular importance.

A very detailed MD investigation of the influence of hydration on the conformations of tripeptides has been carried out by Urbanc and coworkers. They explored the conformational sampling of GxG and AAA in water with different combination of force fields and water models.[68,85,125,126] A comparison of force fields based on these and the works of other research groups can be found in the next section of this article. Here I focus solely on hydration effects. Irrespective of the force fields the results obtained by these authors revealed differences between peptide hydration in pPII and β-strand. Toal *et al.* showed that a reduced hydration of the central alanine residue can explain the slightly lower pPII propensity of the alanine dipeptide compared with trialanine.[85] In another study Meral *et al.* investigated the conformational ensemble of 15 different GxG peptides.[125] They observed that pPII orientations are associated with an increased population of water oriented parallel to the side chain surface (Fig. 14). In contrast, β-strand conformations exhibit more heterogeneous water orientations. These findings suggest that β-strand might be entropically favored over pPII, in full agreement with thermodynamic studies.[69,119,127] A comparison of GAG and AAA by Zhang *et al.* revealed that substituting the two terminal glycines of GAG by alanines leads to an increase of the average number of water molecules as well as of the number of water–water interactions.[68]

While most of the studies performed to elucidate conformational propensities of amino acid residues emphasized the role of hydration exceptions from the rule deserve to be mentioned.[35] Drozdov *et al.* performed Monte Carlo simulations with OPLS parameters and the TIP5P water model to explore how hydration affects the energy landscape of an alanine dipeptide. They arrived at the conclusion that peptide–water interactions favor compact (*i.e.* right-handed helical) rather than extended conformations such as pPII. Preference for the latter is associated with a minimum of the combined torsional and van der Waals interaction energy. In other words: pPII is populated because steric conflicts are avoided. The results of this study are at variance not only with the above

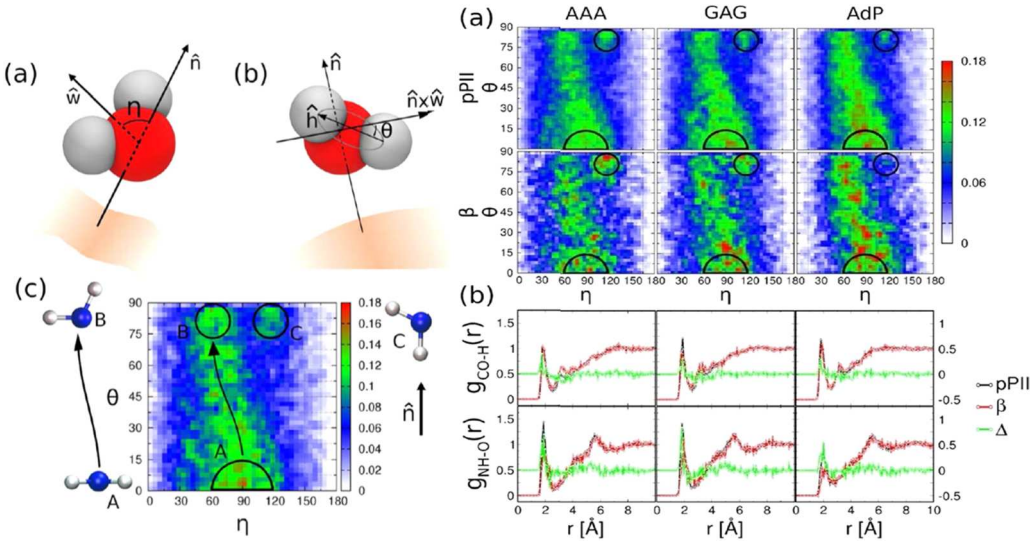

Fig. 14 Left: Illustration of the angles η and θ describing the water orientation in the hydration layer of tripeptides. (a) η is the angle between the normal ($\hat{n}$) on the solvent accessible surface of the peptide and the symmetry axis of water ($\hat{w}$), (b) θ is the rotational angle with respect to this symmetry axes. (c) The water orientation plot for the central A in AAA in the pPII conformation. Regions A, B, and C are outlined alongside the respective water orientations relative to the normal to the SAS of the peptide, $\hat{n}$. Right: Hydration properties of AAA, GAG, and AdP as obtained with the TIP3P water model. (a) Water orientation plots showing distributions of η and θ angles of water surrounding the side chain of (central) A in pPII (top) and β (bottom) conformations. (b) Radial distribution functions of water with respect to CO (top) and NH (bottom) groups of the central A in pPII conformations (black curves), β (red curves) conformations, and the corresponding pPII to β-strand differences (green curves). Reprinted with permission from ref. 125, 2015, American Chemical Society.

cited computational study but also with available experimental data. First of all, they can hardly explain the experimentally established stabilization of β-strand at high temperatures. Second, they contradict the fact that pPII is destabilized in non-aqueous solutions.[128–130]

One might wonder whether the exceptionally high pPII propensity of alanine could be due to the special properties of its methyl side chain. The work of Meral *et al.*[125] seems to support such a view in that it reveals a cage like water structure in the pPII conformation of the alanine residue. However, a *J*-coupling/amide I′ analysis of cationic GGG reveals a pPII propensity of the central glycine residue that is comparable with the one of alanine in GAG (with the pPII fraction equally partitioned between the right and left half of the Ramachandran plot).[131] MD simulations with Amber ff14SB, OPLS-AA and CHATMM36m reproduce this pPII dominance qualitatively, though to a different extent. The result of this study shows that the propensity for pPII is engrained in the backbone and that it is modified by individual side chains of residues. These results might explain the above mentioned observation of 6 pPII helix fold of the crystallized snow flea antifreeze protein which contains 46 glycine residues.[108] Apparently, alanine just stands out because it's methyl group accommodates hydration water as computationally demonstrated.[87,125]

9. Nearest neighbor interactions

The random coil model is based on the assumption that the conformational dynamics of individual residues are uncorrelated. This is generally called the isolated pair hypothesis (IPH).[7] However, multiple lines of evidence gathered over the last 30 years invalidate this assumption. This has far-reaching consequences for our understanding of unfolded/disordered proteins and the thermodynamics of protein folding which have not yet fully recognized and appreciated in the field.

Since this author has recently published a review article summarizing the evidence for nearest neighbor interactions,[132] this section confines itself to a brief summary of experimental results obtained with short peptides. Basically, there are two types of nearest neighbor interactions which ought to be distinguished. In one scenario, it does not matter whether the neighbor adopts pPII, β-strand or turn-supporting conformations; it is just its steric and physicochemical properties that affect the Gibbs energy landscape of a residue. In this case the IPH is not violated because conformational ensembles of residues are still uncorrelated. However, if the interaction energy depends on the conformation of neighbors, the IPH breaks down. As a consequence thermodynamic parameters like conformational enthalpy and entropy and Gibbs solvation energy are no longer additive.[132] The additivity of solvation energies of residues is generally being assumed for estimating the solvation energy contribution to protein folding.[133]

Several studies of nearest neighbor effects in coil libraries have led to the conclusion that particularly aromatic neighbors shift conformational distributions towards β-strand.[95,134] *Per se*, such observations do not allow the identification of the type of nearest neighbor interactions. Avbelj and Baldwin provided some theoretical evidence for the notion that underlying changes of the solvation free energy are indeed

conformation dependent.[135] Sosnick and coworkers constrained MD simulations with coil library information.[136,137] In their thermodynamic model conformational changes between neighbors are correlated, in violation of the IPH. They clearly demonstrated that *e.g.* residual dipole coupling data obtained for apo-myoglobin in 10% acrylamide can only be sufficiently reproduced if conformation dependent nearest neighbor interactions are taken into account.[97] Computational work of Pappu *et al.* suggested some nearest neighbor interaction between residues in helical conformations.[112] Results from Monte Carlo simulations on host–guest peptide systems showed that non-glycine residues populate the upper left quadrant at the expense of right-handed helical conformations.

A systematic and residue specific investigation of nearest neighbor interactions in short peptides has been carried out by Toal *et al.*[138] and more recently by Milorey *et al.*[101,139,140] These authors combined an analysis of *J*-coupling constants and amide I′ profiles to obtain Ramachandran plots of amino acid residues in unblocked tetra- and pentapeptides for different upstream and downstream neighbors. For the sake of brevity, I focus here on the influence of neighbors on alanine and arginine. Alanine is of fundamental importance because of its high abundance in proteins and its particularly high propensity for pPII. Arginine is a frequent contributor to intrinsically disordered segments of proteins (*e.g.* in protoamine sequences).[46] Fig. 15 compares the pPII, β-strand and turn-supporting fractions of alanine in various tri- and tetrapeptides.[141] The turn-supporting fraction encompasses all residue conformations that appear in β-, γ- and asx-turns. While alanine neighbors slightly increase the pPII populations, any of the investigated non-alanine neighbors stabilizes β-strand over pPII. Nearest neighbor interactions are particularly significant for GSAG and GAVG. A comparison of arginine mole fractions in GRG, GRRG and GRRRG is shown in Fig. 16.[139] The data reveal that R2 is particularly affected by nearest neighbor interactions which substantially stabilize β-strand over pPII. Milorey *et al.* showed that based on their results the end to end distance of a statistical coil of a poly-L-arginine peptide would be more extended than a self-avoiding random coil.

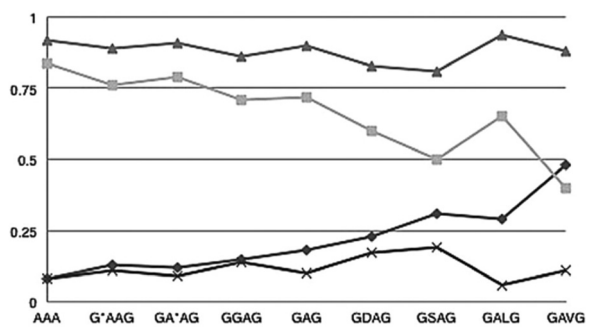

Fig. 15 Mole fractions of pPII (grey squares), β-strand (black rhombus), and turn-supporting conformations (cross) of alanine in the indicated tri- and tetrapeptides. The black triangle data points represent the sum of pPII and β-strand populations. Reprinted with permission from ref. 138, 2015, Wiley & Sons.

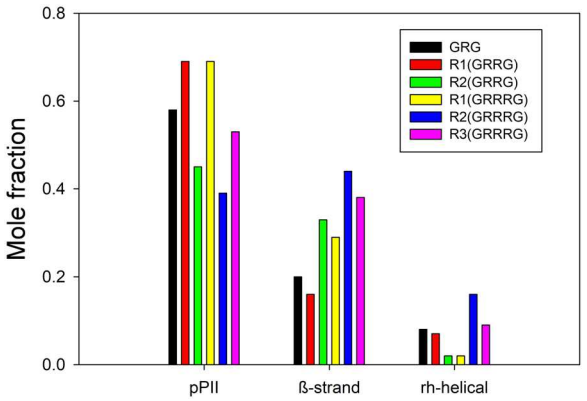

Fig. 16 Mole fractions of pPII, β-strand and right-handed helical conformations of arginine in the indicated peptides. The data were taken from ref. 139.

Thus far I discussed only nearest neighbor induced changes of conformational propensities. However, as shown in the above cited papers they can affect the positions of the basin centers as well. For alanine, serine and leucine as neighbors decrease the ψ-values of both pPII and β-strand. Valine as upstream neighbor shifts both basins to the left. For the R-containing peptides, basin shifts are even more pronounced for the φ-coordinate of the β-strand basin which is moved substantially to the left. pPII and β-strand now appear very clearly separated in the Ramachandran plot (Fig. 17). Toal *et al.* reported a similar effect for leucine in GLyG and GxLG peptides (x and y denote different guest residues).[138] Interestingly, the underlying nearest neighbor interactions do not have a significant influence of on the conformational propensities of leucine. As shown by Schweitzer-Stenner and Toal shifts of basin coordinates can make Ramachandran distributions more dissimilar than even rather significant changes of conformational propensities.[142]

The influence of nearest neighbors on the position of basins in the Ramachandran space described above is significant. This observation leads to the conclusion that the use of coil library distributions that average over the influence of nearest neighbors are of limited usability for the structural analysis of guest residues in glycine based host–guest systems like G_2xG_2 and blocked dipeptides.[69,73,143] Any attempt in this regard should be exclusively based on experimental data obtained for the investigated peptide as exemplified by the combined use of complementary spectroscopic methods.[66,90,91]

Meta analyses of the above data provided strong evidence for the notion that the nearest-neighbor interactions are predominantly governed by pPII-β interactions which can be cooperative or anti-cooperative. For the cases discussed above, the interaction is cooperative, *i.e.* pPII-β sequences are stabilized over pPII–pPII and β–β sequences.[110,140] The results of these analyses have important implications. On a first glance nearest neighbors seem to randomize distributions, *i.e.* increasing the conformational entropy and making the Ramachandran distributions more 'random coil' like. However, such a picture

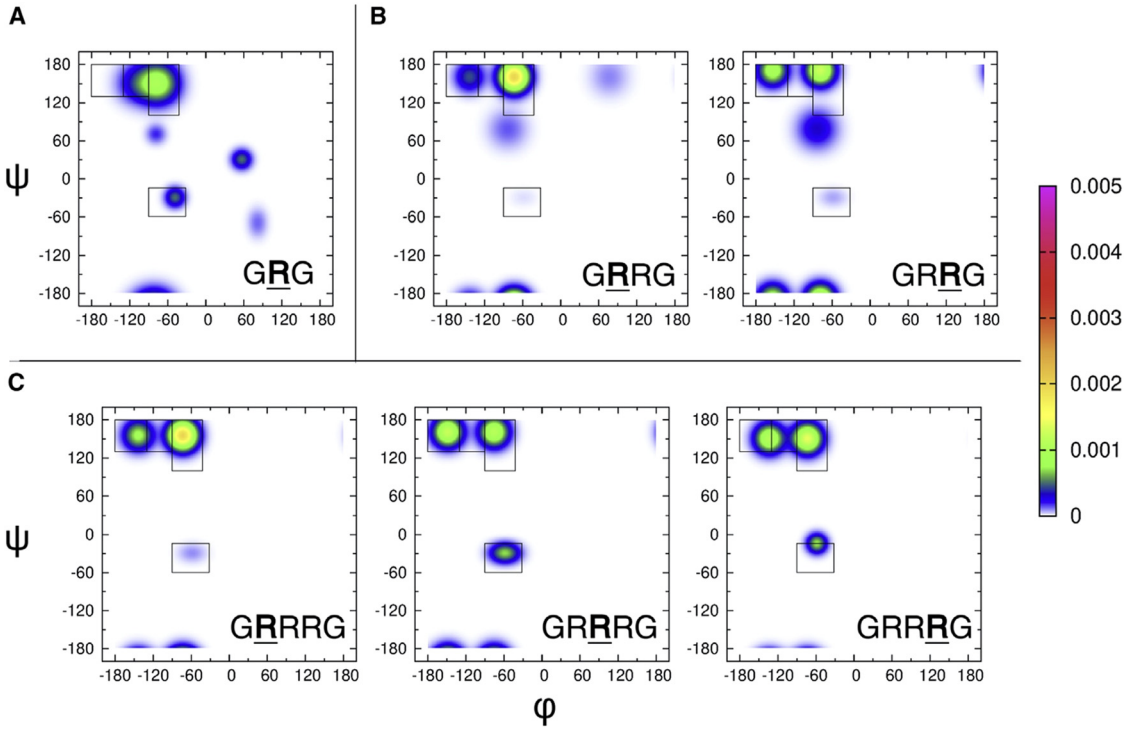

Fig. 17 Ramachandran plots of arginine residues in the indicated tetra- and pentapeptides obtained from a global analysis of *J*-coupling constants and amide I′ profiles. Reprinted with permission from ref. 139, 2021, Elsevier.

could be misleading since individual Ramachandran plots do not tell us anything about inter-residue correlations which as in the above systems can reduce the conformational entropy of the entire peptide. For the guest residues of GxyG peptides the reduction of the entropy contribution to the Gibbs energy generally lies in the 10^2 J mol^{-1} at room temperature, but for some xy pairs (SV and FD) respective values are larger than 1 kJ mol^{-1}. For the homopeptides GRRRG and GDDDG the entropy reduction is significantly more pronounced (2–4 kJ mol^{-1}).[140]

10. Role of short unfolded peptides for MD force field assessment and development

Among the many purposes served by obtaining reliable experiment-based data about conformational propensities of amino acid residues is to facilitate the development of molecular dynamics force fields. Despite of many advancements in past two decades, MD force fields are still not accurate enough to fully capture dynamics of unfolded and intrinsically disordered proteins. Since the early days of the discovery of alanine' pPII propensity attempts have been made to develop force fields that are in line with experimental data. Gnanakaran and Garcia made a brute force attempt in that they eliminated the torsional force constants for φ and ψ for an Amber 94 force field to obtain very high pPII propensities for alanine in oligo-L-alanine peptides.[116] Less radical strategies have been pursued since then, with very mixed results. In order to keep this section brief I focus on attempts guided by published *J*-coupling constants.

In what follows a yet not fully solved problem must be briefly discussed. Attempts based on optimizing force fields have been focused on alanine owing to the availability of *J*-coupling constants for a variety of oligo-alanine peptides from the work of Graf *et al.*[66] In order to demonstrate the quality of force field improvement deviations between calculated and experimental *J*-coupling constants relied on the reduced χ^2-function:

$$\chi_R{}^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{\left(J_{i,\mathrm{exp}} - \langle J_{i,\mathrm{calc}} \rangle\right)^2}{\sigma_i^2} \qquad (2)$$

where N is the number of considered *J*-coupling constants, $J_{i,\mathrm{exp}}$ are the individual experimental coupling constant values, $\langle J_i \rangle_{\mathrm{calc}}$ are the values calculated for the final Ramachandran distribution of a residue and σ_i the statistical error of the coupling constants. The use of eqn (2) would be straightforward if reliable values were available for the latter. Generally, one would associate such statistical errors with the experimental data. In the case of *J*-coupling constant, the respective values are generally small and would therefore allow for a high precision assessment of calculated coupling constants. Unfortunately, the main contribution to σ_i is associated with the calculated values because they depend on the accuracy of the Karplus parameters in eqn (1). The amplitudes A, B, C and the phases θ_i are empirical parameters that researchers have obtained from fits to *J*-coupling constants observed for proteins for which high quality crystal structures or NMR-based structures are available. Scattering of respective data sets can be

considerable, which leads to uncertainties of the empirical parameters. Bax and coworkers derived error estimates for most of the coupling constants used by Graf *et al.*[66] from X-ray and NMR structure data which seem to be an appropriate choice.[56,57] However, in many studies σ_i was estimated from the scattering of the data to which Karplus equation had been fitted (*vide infra*). These values are often unreasonable large (*e.g.* ±0.5 Hz for $^3J(H^NH^{C\alpha})$), which in turn makes significant deviation between experimental and computed coupling constant looking satisfactory. The main problem with all these approaches is that a set of coupling constants obtained either from X-ray or NMR data set is not entirely a statistical ensemble. There is no doubt that part of the scattering is statistical in nature (uncertainties of *J*-coupling measurements for large system like proteins and uncertainties of dihedral angles) but an equal or even dominant part could reflect residue specific deviations which remain unspecified. This means that for each residue the error should be in part systematic in nature.

Computational chemists have tried to address the later issue by using DFT calculations to determine the Karplus parameters for alanine.[67] Fig. 6 compares Karplus curves calculated with different empirical and DFT-based parameters (for alanine). Despite the differences between the empirical parameters the corresponding Karplus curves are very similar, with the exceptions of the region around $\varphi = 120^\circ$ for $^3J(H^NC_\beta)$ that lies in the forbidden region of the Ramachandran plot. DFT based Karplus curves are more pronounced at extrema for $^3J(H^NH^{C\alpha})$, $^3J(H^NC')$ and particularly for the -60° region of $^3J(H^NC_\beta)$. Some of these discrepancies can be explained by the avoidance of dynamic averaging in DFT calculations,[67] but it is difficult to explain the behavior of $^3J(H^NC_\beta)$ which puts values way outside of any measured experimental value.

Best *et al.* used two modifications of the Amber force field termed ff99SB and ff03w to reproduce the *J*-coupling constants that Graf *et al.* reported for penta-alanine.[144,145] The force field modifications were based on quantum chemical calculations. The theoretical coupling constants were calculated by means of the Karplus equation with DFT based parameters (DFT2). The conformational distributions obtained with the original and the modified force fields were significantly different from the ones reported by Graf *et al.*[66] and Toal *et al.*[85] For none of the obtained Ramachandran plots did the pPII fraction exceed 0.5. The obtained distributions are nearly random-coil like. While the authors reported convincing χ_r^2-values (all below 2), they did not provide a direct listing of the computed coupling constants. In a response to this work Verbaro *et al.* used the IR and VCD amide I′ profiles and end to end distance measurements with fluorescence resonance energy transfer to show that the results Best and Hummer reported for A_5 do not capture the properties of A_5W, which were found to sample significantly more extended structures.[146] Not surprisingly, an analysis of the spectroscopic data, which included the *J*-coupling constants of Graf *et al.*, yielded a much higher pPII content.

In parallel to the above cited work Nerenberg and Head-Gordon used the AMBER ff99SB forcefield in conjunction with TIP3P and newer water model TIP4P-Ew to produce conformational ensembles of cationic AAA, GGG and VVV for different temperatures.[147] As Best and coworkers, they utilized DFT Karplus equation for their calculation of ensemble averaged coupling constants in addition to the empirical parameters of Hu and Bax.[57] For A_3 they found a slight stabilization of pPII in TIP4P-Ew water (compared with TIP3P). No specifics about conformational distributions were provided for GGG and VVV. The authors confined themselves on comparing reduced chi-square values (eqn (1)). For simulation at room temperature, they obtained χ_r^2-values between 1.95 and 2.92 if the coupling constants were calculated with empirical Karplus parameters. Comparatively low values were obtained with both DFT-based parameters sets. For G_3, the χ_R^2-values were high for all Karplus parameter sets, but a substantial improvement was obtained if $^3J(C'C')$ and $^2J(N'C_\alpha)$ were not considered. For V_3, the omission of $^3J(C'C')$ led to a substantial reduction of χ_R^2 (*i.e.* below 2) for the empirical set and DFT2. In addition, the authors performed replica exchange MD simulations with TIP4P-Ew for five GxG peptides. To improve the agreement with the experimental data of Hagarman *et al.*,[91] the authors reduced the $n = 2$ potential term in the respective expressions for dihedral angles. Here, only the Hu and Bax Karplus parameters[57] were used. The authors reported quite satisfactory χ_R^2-values, but unfortunately no specifics about conformational distributions.

One of the latest revisions of Amber force fields was carried out by Tian *et al.*[149] Their force field termed Amber ff19SB. The authors obtained torsional φ/ψ-parameters by means of fits to energy surfaces obtained from DFT calculations for different amino acid residues (glycine, alanine, valine, leucine). The latter were carried out in an implicit solvent.

Fig. 18 compares the Ramachandran plots for the alanine and valine dipeptide obtained with Amber ff14SB and Amber ff19SB. The distributions obtained with the former are hardly distinguishable, both a dominated by pPII, at variance with experimental data (*vide supra*). With ff19SB, however, clear differences emerge, *i.e.* a redistribution of sampling from pPII to β-strand for valine. A similar result was obtained for leucine. This means that ff19SB at least accounts for residue specific conformational propensities. The authors used $^3J(H^NH^{C\alpha})$ parameter of 19 dipeptides (proline excluded) reported by Avbelj *et al.*[143] to compare ff14SB and ff19SB further. The results are actually mixed. The latter performed better for some residue (besides V and L), for protonated H, C, N and protonated K, but ff14SB yielded a better fit for quite a large set of residues.

Over the last 10 years Urbanc and colleagues undertook a systematic investigation of different force field – water model combinations. Contrary to most of the work described above they used unblocked GxG peptides as benchmark systems. Thus, they could take advantage of a much larger set of experimental data, *i.e.* five different *J*-coupling constants as well as amide I′ profiles. They compared the performance of the investigated force fields with the one achieved by the above cited Gaussian model of Schweitzer-Stenner.[81] The main results of their works can be summarized as follows. First, for alanine (GAG and AAA) none of the investigated force fields (Amber

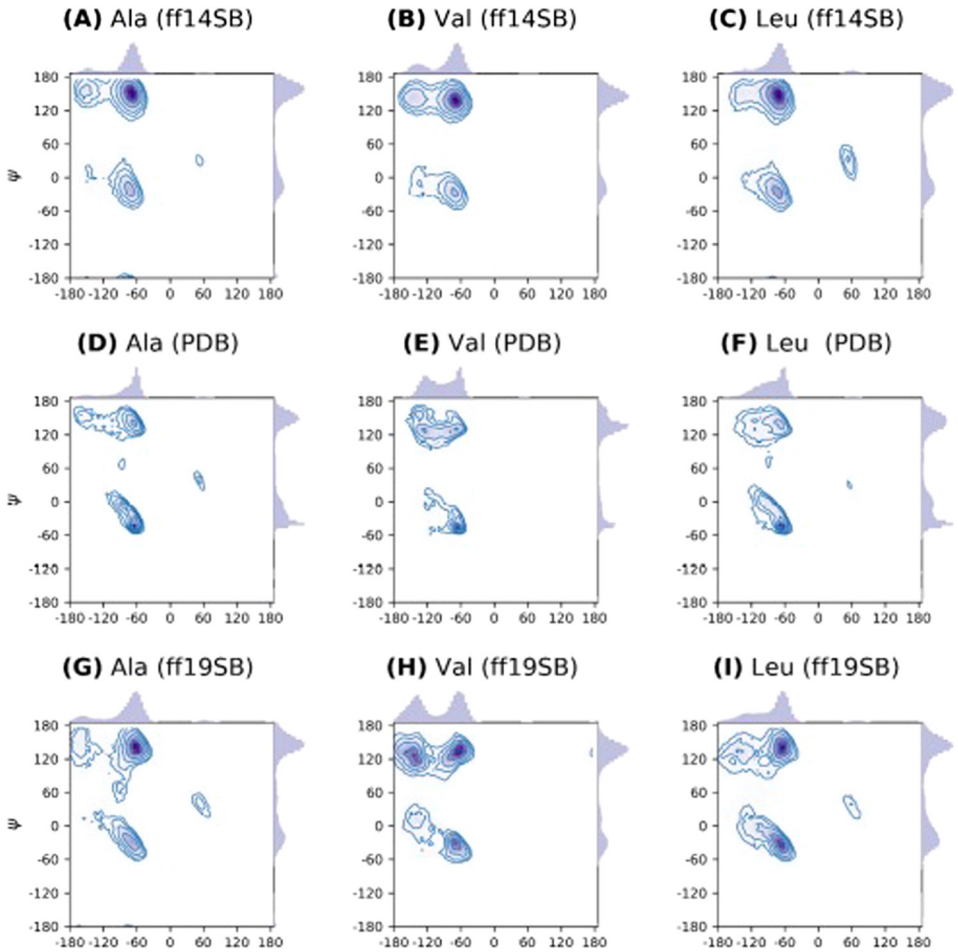

Fig. 18 Ramachandran plots of alanine, valine and leucine dipeptides obtained from MD simulations with Amber ff14SB + OPC water (upper panel), a coil library (middle panel, data from Lovell *et al.*[148]) and MD simulations with Amber ff19SB + OPC water. Each contour line represents a doubling in population. Reprinted with permission from ref. 149, 2020, American Chemical Society.

ff14SB, ff99SBnmr, ff03ws, OPLS-AA/L, OPLS-AA/M and CHARMM36) reproduces experimental data as well as the Gaussian model.[68] Second, among these force fields ff14SB with TIP3P water produces the best *J*-coupling constants and amide I′ profiles. Fig. 19 compares the Ramachandran plots obtained with the above force fields. Third, regarding the *J*-coupling constants, CHARMM36m produces results close to the one obtained with the Gaussian model. Fourth, in another study from this laboratory, OPLS-AA/M, CHARMM36m, Amber ff14SB and the more novel Amber ff19SB were used to produce Ramachandran plots of various GxG peptides for which *J*-coupling constants and amide I′ profiles had been reported.[126] None of the force field performed satisfactorily. However, ff19SB at least captured a trend displayed by the varying pPII propensities of the investigated amino acid residues. Since the other force fields were optimized for alanine, they failed to account for these differences.

Finally, we mention an approach to MD simulations that differs conceptually from the ones discussed above. Rather than dealing with explicit water models, Vitalis and Pappu developed a continuous solvation model.[150] Here, the transfer from the gas phase to solution is accounted for by direct mean field interactions and the screening of interactions between polar groups. They combined this solvation model with modified versions of classical force fields where the torsional potentials were omitted. In that regard their work resembles the one of Gnanakaran and Garcia.[86] As others, they used the $^{3}J(H^{N}H^{C\alpha})$ of dipeptides for validation. Irrespective of the utilized force field the calculated *J*-coupling constants all cluster in the region between 7 and 7.5 Hz, which is clearly at variance with the experimental data. The authors gained some confidence in their model by a comparison with earlier results of DFT calculations. However, it is unclear how this can reconcile the discrepancy between theory and experiment.

A very radical approach by Elcock and coworkers deserves to be mentioned. They produced various extensions of the Amber ff99SB force field with an increasing number of modifications to capture side chain specifics.[151,152] The final version termed RSFF2 were found to reproduce $^{3}J(H^{N}H^{C\alpha})$ coupling parameters of blocked tripeptides reported by Cho and coworkers.[153] The authors judged the suitability of their force field solely by regression coefficients obtained from correlation plots of calculated and experimental *J*-coupling constant. In order to gain more credibility, this force field should be applied to much larger sets of *J*-coupling constants described above.

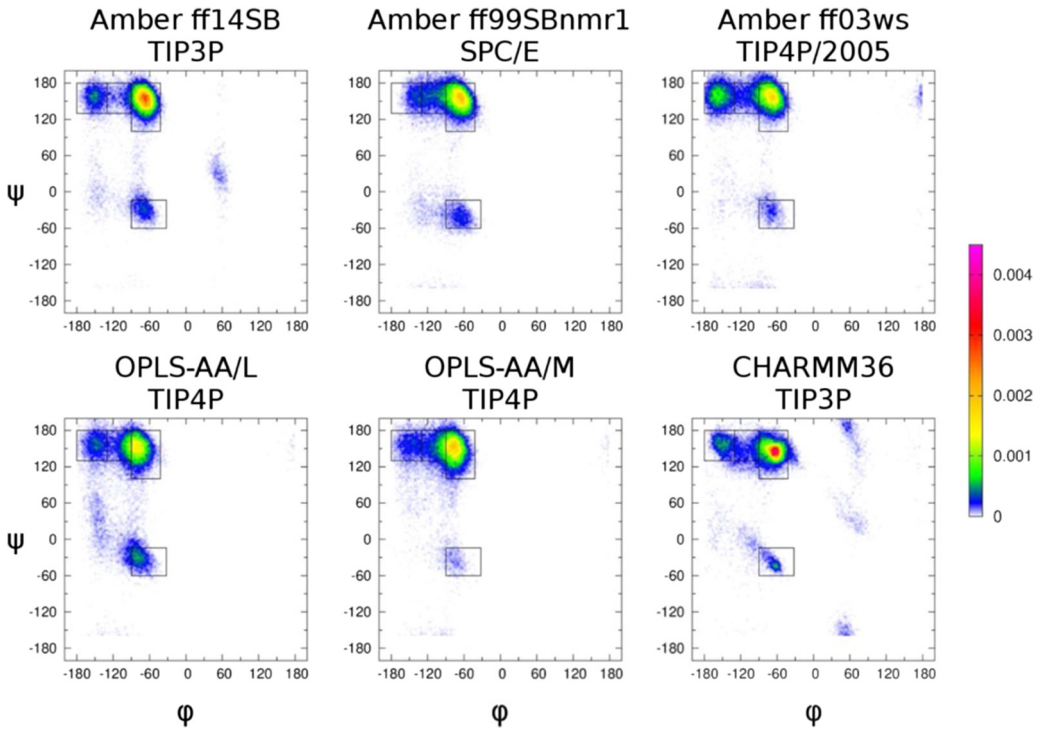

Fig. 19 Ramachandran plots of cationic GAG obtained from MD simulations with the indicated force field and water model combination. The rectangles are mesostates associated with pPII, β-strand and right-handed helical conformations. Reprinted with permission from ref. 68, 2020, American Chemical Society.

Very recently, Yuan and Wang reported a DFT based analysis of the dialanine peptide and zwitterionic unblocked oligoalanine peptides for which they considered hydration implicitly by employing the conductor-like screening model.[154] They constructed a Born–Oppenheimer type energy surface by using the so-called adaptive force matching method. Based on the reported χ_R^2-values they managed to account for the *J*-coupling constant values of Graf *et al.*[66] but computed coupling constants were not explicitly reported. It is remarkable, however, that the obtained force field produced pPII fractions above 0.9 which is good agreement with results obtained with the Gaussian model.[68,85] From a comparison of force fields obtained with and without the above implicit solvent model the authors arrived at the conclusion that pPII and right handed helical conformations are stabilized by solvent polarization over β-strand conformations. Apparently, the developed force field is residue specific. It remains to be seen whether similar strategies lead to a better understanding of different residue propensities and nearest-neighbor interactions.

Other approaches aimed at force field development utilized coil library distributions. Since this article focusses on the use of short peptides, we just refer the interested reader to the relevant literature.[155–158]

11. The usability of short peptides: assessment and outlook

Thus far this article has provided mostly a critical overview of how short peptides have been used to determine residue specific propensities for backbone conformations and their dependence on solvation and nearest neighbors. In this concluding chapter I briefly discuss to what extent short peptides are and could be used as reference systems for the investigation of IDPs and unfolded proteins. To this end I focus current structural analyses of IDPs by NMR spectroscopy, the relevance of nearest neighbor interactions, the conformational dynamics of side chains and force field development.

11.1 Short peptides as reference systems for the use of secondary chemical shifts

Are short peptides useful for developing an understanding of unfolded proteins and intrinsically disordered proteins? One might be doubtful about this idea, because the environment of amino acid residues in water might not be identical with the one in unfolded/denatured systems for which the scaling exponent lies below 0.5. Proteins denatured in good solvents (let's assume for a moment that classical denaturing reagents fall into this category) seem to show similar scaling laws irrespective of their amino acid residue composition. For IDPs with a high net charge global parameters like the radius of gyration or the end to end distance seem to be describable solely by their charge balance.[46]

There are several lines of arguments in favor of short peptide investigations. First, from a physical chemistry point of view, they are ideal systems to study the interplay between backbone, side chain and solvent the detailed knowledge of which is crucial for an understanding of unfolded states and of folding/unfolding processes. Blocked dipeptides and unblocked GxG typed tripeptides are suitable tools to explore the intrinsic

propensities of amino acid residues. Detailed and reliable information about the latter allow the construction of a reference system based on which any additional interactions in more complex molecular environments can be determined. Second, and this is a corollary of the first argument, they are ideal benchmark systems to assess the quality of molecular dynamics force fields that are used to model the behavior of unfolded/intrinsically disordered peptides. Third, knowing intrinsic propensities of amino acid residues can be helpful to calculate the entropy of unfolded/disordered systems, which is important for an understanding of disorder to order transitions of all kinds.[97]

Besides being directly relevant for the analysis of unfolded/disordered proteins the results of the above reviewed work on short peptides should be also of importance for the NMR-based structure analyses of these systems. In addition to *J*-coupling constants chemical shifts (CS) of ^{1}H, ^{13}C and ^{15}N nuclei are frequently used for the analyses of their structures. To this end, CS values of guest residues in short model glycine base peptides are used as reference values which are thought to represent the (local) random coil state of these residues.[41,159,160] Hence, any statistically significant deviation from these CS values is interpreted as indicating local deviations from the random coil distribution. These changes are termed secondary chemical shift. Even if one uses the more appropriate term statistical coil the concept underlying this approach looks convincing, since by utilizing individual values for each residue differences between their Ramachandran distributions are automatically taken into account. However, two issue remain unresolved. First, the CS depends on the environment. Differences between *e.g.* pH and temperature can be accounted for by measuring the CS of the reference peptides as a function of these parameters.[160,161] However, at least the CS of ^{1}H have been shown to depend on the solvent exposure of the respective functional group.[162,163] Therefore, the CS can change if a residue is moved from the (good) solvent into the hydrophobic interior of a collapsed but unfolded protein. This change is likely to reflect changes of the respective conformational distributions as well as intrinsic electronic effects. Second, the CS can be expected to be sensitive to nearest neighbor interactions. These interactions involve physical effects (shielding and deshielding) and structural dependencies as discussed in this article. The influence of nearest neighbor ion chemical shifts has been recognized at a very early stage. Attempts to quantify such effects involved the use of *e.g.* unblocked GGxA peptides where the influence of the guest residue x on the CS on alanine was determined.[161,164] Another approach used GGxGG peptides to obtain the influence of x on glycine.[165] A more recent attempt utilized QQxQQ,[166] since Q was thought to be more representative of amino acid residues. In view of the above results that emerged from studies on tetra- and pentapeptides it seems to be questionable whether the structural part of nearest neighbor interactions is sufficiently represented by these peptides. A much broader approach has recently been taken that utilized the distribution of chemical shifts in the spectra of intrinsically disordered proteins.[167] It would be of interest to explore whether the obtained result show any correlation with the nearest neighbor interactions in short peptides.

11.2 Nearest neighbor interactions

The above cited NMR studies emphasize the relevance of nearest neighbor interactions. This issue, however, has thus far been only incompletely addressed. If, as results have shown, these interactions depend on residue conformations, they invalidate the isolated pair hypothesis. Strictly speaking this means that the use of the term random coil is not permissible irrespective of a protein's global behavior. Attempts have been made to infer nearest neighbor interactions from coil libraries[96,98,135,136,155] but specific information about the underlying mechanism and residue specificity have not emerged from these studies. It should be mentioned, however, that Sosnick and coworkers demonstrated their relevance for unfolded proteins.[97,137] The first attempt to specifically investigate nearest neighbor interactions by Toal *et al.* and Milorey *et al.* provided some useful information about the nature of residue pairs and their respective conformations determine nearest neighbor interactions.[101,138,139,140] To my best knowledge these works are the only ones that were based on a sufficiently large set of experimental data that allowed for the construction of Ramachandran plots and a quantitative assessment of nearest neighbor interactions. Schweitzer-Stenner and Toal provided evidence for their applicability to rather large denatured and intrinsically disordered proteins. However, owing to the amount of work that must be invested in determining nearest neighbor interactions the data set is still rather limited. Extended it to all combinations of residues with upstream and downstream neighbors is out of question. It would make sense instead to continue the above experimental work with representatives of different residue groups (*i.e.* aliphatic, aromatic, dipolar, ionized). Nearest neighbor interactions can significantly change the conformational entropy.[97,140] What this means for an entire protein has still to be explored. Generally, investigating nearest neighbor interactions has run out of steam, most likely because their explicit consideration would significantly increase the complexity of models for unfolded and disordered proteins.

Nearest neighbor interactions discussed thus far generally do not account for the local formation of hydrogen bonding. The above work on tri-, tetra and pentapeptides yielded evidence for the population of turn-supporting conformations which require the interpeptide hydrogen bonding.[90,93,101] However, the examination of the formation of *e.g.* classical β-turns would require longer peptides.[168] It is noteworthy in this context that the Zimm–Bragg as well as the Lifson–Roig theory predict an increase of helical content with increasing peptide length.[169,170] For alanine, which exhibits the highest propensity for right-handed helices,[171,172] This notion is corroborated by MD simulations of oligo-alanines and experimental data.[86,146] Obviously, any theoretical approach has to go beyond nearest neighbor interactions for longer oligopeptides and certainly also for unfolded and disordered proteins.

11.3 Structural heterogeneity of side chains

An additional complication which has not even been addressed in this article arise from the structural heterogeneity of side chains which can populate different rotamers regarding different dihedral angles χ_i, i = 1–4, depending on the length of the side chain (Fig. 20). For the sake of brevity, I confine myself on discussing solely χ_1. The three rotamers regarding this angle are illustrated in the left part of Fig. 20. Analyses of coil libraries and MD simulations have clearly revealed that different amino acid side chains differ in terms of the population of these rotamers and that these populations depends on the backbone conformation.[155,174–177] Experimental work exploring the rotamer populations of residues in short peptide is practically non existing. The only exception I am aware of is the paper of Rybka *et al.*, who measured $^3J(H^{C\alpha}H^{C\beta})$ coupling constant of the two C_β-protons to determine the rotamer distributions of GNG, protonated and ionized GDG and of the protonated blocked tetrapeptide Ac-GDG. In all these cases the -60° rotamer was found to be the most populated one with mole fraction ranging from 0.51 for Ac-GDG and 0.74 for ionized GDG. The substantial population of the 180° was found to be consistent with a significant sampling of asx-turns by protonated D. While the 60° rotamer was found to be preferred in pPII and β-strand conformations, the -60° rotamer coexists with turn II′ β_{i+2} conformations which have been shown to be disproportional populated by protonated and ionized D-residues.[93,101,178] In an earlier reported analysis of coil libraries Jiang *et al.* found that D and N combined show a preference for pPII, a combined region of right handed helical and turn II′ β_{i+2} and for left handed helical conformation for the g^+ (χ_1 = 60° and 180°) conformation. On the contrary, the Ramachandran plot for *t* conformations ($\chi_1 = -60^\circ$ and 180°) is rather peculiar in that it is mostly populated in a region comprising pPII and type II β_{i+1} turn conformations. The coil library distributions show some similarities with the ones obtained for GDG and GNG, but differences are also noteworthy. This is not surprising since D is significantly affected by it nearest neighbors.[139,141,178]

There is no doubt that a complete understand of conformational preferences of amino acid residues in unfolded and disordered proteins requires a more thorough analysis of the relationship between side chain and backbone conformations. Short peptides and NMR spectroscopy would be ideally suited for this goal. The necessity to explore the mutual dependence of side chain and backbone conformations has been recognized by Sosnick and coworkers.[179] They showed the folding of ubiquitin is associated with a side chain entropy loss that contributes 20% to the overall decrease in conformational entropy. The question arises to what extent backbone dependent conformational propensities of side chains play a role in nearest neighbor interactions. Exploring the interplay between backbone conformations, side chain rotamer populations and hydration is a still to be carried out project the results of which will be essential for an understanding of unfolded and disordered proteins and peptides.

11.4 Force fields for IDPs

Obviously, experimentally determined conformational propensities of amino acids are ideally suited for developing molecular dynamics force fields. The success claimed by some researchers was mostly built on very limited data sets and/or rather generous assessments of statistical errors of coupling constant (the larger the error the better even insufficient reproductions of experimental data). The work of the Urbanc group has clearly shown that currently none of the already optimized force fields is capable to sufficiently reproduce the *J*-coupling constants and amide I profiles of GxG peptides. Only the new Amber ff19SB does at least capture differences between amino acid residues with regard to their pPII propensities.[126,180] It seems to be obvious that optimizing a force field for one amino acid residue (alanine) produces poor results for other residues. The situation becomes worse for longer peptides where current force fields are incapable of catching nearest neighbor interactions. This is a serious and thus far underestimated problem, which is likely to reflect the insufficient description of cooperativity between the hydration shells of residues. Generally, one should be skeptical about attempts to develop force field based on DFT calculations in implicit water. Work of Wong, Lanza and their respective coworkers have clearly show that implicit water cannot catch conformational propensities the way explicit water does.[121,124] Even though computational expensive, calculations with explicit water will be necessary to obtain reliable energy surfaces in the Ramachandran space. The role of the solvent in nearest neighbor interactions has been demonstrated by Toal *et al.*, who showed that even in the absence of strong interactions at room temperature entropic effects can completely change the picture at temperatures at

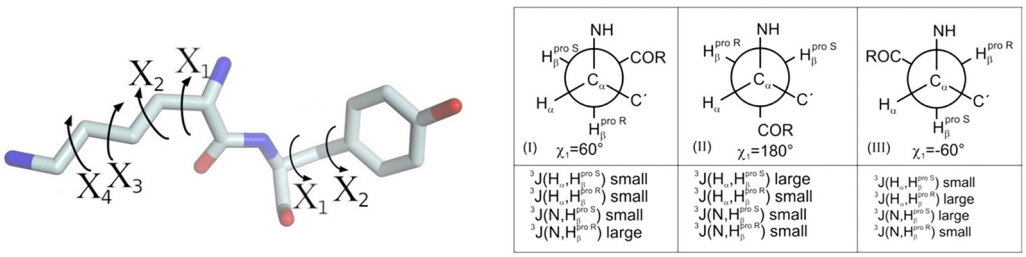

Fig. 20 Left: Illustration of dihedral rotamer angles of the blocked tripeptide lysyltyrosine.[173] Right: Newman projection of the three rotamers associated with the dihedral angle χ_1 ($(C'-C_\alpha-C_\beta-H_{\beta_i}$ i = 1, 2) for pro *R* and pro *S*, respectively; 60°: g^-, g^+ for *R* and *S*, 180°: g^+, *t* for *R* and *S*, -60°, *t*, g^- for *R* and *S*; *g*: *gauche*, *t*: *trans*). Reprinted with permission from ref. 93, 2013, Wiley & Sons.

which many proteins melt.[138] The physical reasons for these observations remain elusive.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

I would like to thank Prof. Brigita Urbanc (Department of Physics, Drexel University) and Dr Siobhan Toal (Department of Chemistry and Biochemistry, Rowan University) for a thorough and critical reading of the manuscript. Work of my own research group covered in this article has been carried out by Drs Andrew Hagarman, Siobhan Toal, Bridget Milorey and Thomas Measey when they were graduate students in our Biospectroscopy Research Group. Their work was performed in close collaboration with Prof. Harald Schwalbe at the Johann Wolfgang Goethe University in Frankfurt, Germany. It has been supported by the National Science Foundation (Chem 0804492 and MCB-1817650).

References

1 P. Almeida, *Proteins: Concepts in Biochemistry*, Garland Science, 2016.
2 J. Herzfeld and H. E. Stanley, A general approach to co-operativity and its application to the oxygen equilibrium of hemoglobin and its effectors, *J. Mol. Biol.*, 1974, **82**, 231–265.
3 M. Antonini and E. Brunori, Hemoglobin and myoglobin in their reactions with ligands, in *Frontiers of Biology*, Elsevier, Amsterdam, 1971.
4 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, Stereochemistry of polypeptide chain configurations, *J. Mol. Biol.*, 1963, **7**, 95–99.
5 D. A. Brant and P. J. Flory, The Configuration of Random Polypeptide Chains. I. Experimental Results, *J. Am. Chem. Soc.*, 1965, **87**, 2788–2791.
6 D. A. Brant and P. J. Flory, The Configuration of Random Polypeptide Chains. II. Theory, *J. Am. Chem. Soc.*, 1965, **87**, 2791–2800.
7 P. J. Flory, *Statistical Mechanics of Chain Molecules*, Cornell University Press, Ithaca, 1953.
8 A. K. Dunker and Z. Obradovic, The protein trinity – Linking function and disorder, *Nat. Biotechnol.*, 2001, **19**, 805–806.
9 A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, Flexible nets: the roles of intrinsic disorder in protein interaction networks, *FEBS J.*, 2005, **272**, 5129–5148.
10 V. N. Uversky, Intrinsically Disordered Proteins, *Brenner's Encycl. Genet.*, 2nd edn, 2013, **vol. 19**, pp. 124–126.
11 V. N. Uversky and A. K. Dunker, *Protein Pept. Folding, Misfolding, Non-Folding*, 2012, pp. 1–54.
12 C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky and A. K. Dunker, Flexible nets: disorder and induced fit in the association of p53 and 14-3-3 with their partners, *BMC Genomics*, 2008, **9**, 1–20.
13 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker and V. N. Uversky, Analysis of Molecular Recognition Features (MoRFs), *J. Mol. Biol.*, 2006, **362**, 1043–1059.
14 N. E. Davey, J. L. Cowan, D. C. Shields, T. J. Gibson, M. J. Coldwell and R. J. Edwards, SLIMPrints: conservation-based discovery of funcitonal motif fingerprints in intrinsically disordered protein regions, *Nucleic Acids Res.*, 2012, **40**, 10628–10641.
15 V. Neduva and R. B. Russell, DILIMOT: discovery of linear motifs in proteins, *Nucleic Acids Res.*, 2006, **34**, W350–W355.
16 V. Neduva and R. B. Russell, Linear motifs: evolutionary interaction switches, *FEBS Lett.*, 2005, **579**, 3342–3345.
17 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life, *J. Mol. Biol.*, 2004, **337**, 635–645.
18 C. J. Oldfield, B. Cue, A. K. Dunker and V. N. Uversky, in *Protein and Peptide Folding, Misfolding and Non-Folding*, ed. R. Schweitzer-Stenner, Wiley & Sons, Inc, Hoboken, 2012, pp. 241–278.
19 J. A. Drake and B. M. Pettitt, Thermodynamics of Conformational Transitions in a Disordered Protein Backbone Model, *Biophys. J.*, 2018, **114**, 2799–2810.
20 J. A. Drake, J. A. Drake and B. M. Pettitt, Physical Chemistry of the Protein Backbone: Enabling the Mechanisms of Intrinsic Protein Disorder, *J. Phys. Chem. B*, 2020, **124**, 4379–4390.
21 A. Avni, H. M. Swasthi, A. Majumdar and S. Mukhopadhyay, Intrinsically disordered proteins in the formation of functional amyloids from bacteria to humans, *Prog. Mol. Biol. Transl. Sci.*, 2019, **166**, 109–143.
22 V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic and A. K. Keith, Unfoldomics of human diseases: linking protein intrinsic disorder with diseases, *BMC Genomics*, 2009, **10**, S1–S7.
23 S. Mukhopadhyay, The dynamism of intrinsically disordered proteins: binding-induced folding, amyloid formation, and phase separation, *J. Phys. Chem. B*, 2020, **124**, 11541–11560.
24 J. Bürgi, B. Xue, V. N. Uversky and F. G. Van Der Goot, Intrinsic disorder in transmembrane proteins: roles in signaling and topology prediction, *PLoS One*, 2016, **11**(7), e0158594.
25 J. Hermans, The amino acid dipeptide: small but still influential after 50 years, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3095–3096.
26 Ramachandran Plot, **https://en.wikipedia.org/wiki/Ramachandran_plot#/media/File:Ramachandran_plot_original_outlines.jpg**.
27 V. N. Uversky, Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature,

pH, Counter ions, membranes, binding partners, osmolytes, and macromolecular crowding, *Protein J.*, 2009, **28**, 305–325.
28 G. N. Ramachandran and V. Sasisekharan, Conformation of Polypeptides and Proteins, *Adv. Protein Chem.*, 1968, **23**, 283–437.
29 C. Ramakrishnan and G. N. Ramachandran, Stereochemical Criteria for Polypeptide and Protein Chain Conformation. II. Allowed Conformations for a Pair of Peptide Units, *Biophys. J.*, 1965, **5**, 909–933.
30 M. Feig, Is alanine dipeptide a good model for representing the torsional preferences of protein backbones?, *J. Chem. Theory Comput.*, 2008, **4**, 1555–1564.
31 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
32 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
33 M. P. Gaigeot, Infrared spectroscopy of the alanine dipeptide analog in liquid water with DFT-MD. Direct evidence for PII/β conformations, *Phys. Chem. Chem. Phys.*, 2010, **12**, 10198–10209.
34 H. Okumura and Y. Okamoto, Temperature and pressure dependence of alanine dipeptide studied by multibaric-multithermal molecular dynamics simulations, *J. Phys. Chem. B*, 2008, **112**, 12038–12049.
35 A. N. Drozdov, A. Grossfield and R. V. Pappu, Role of Solvent in Determining Conformational Preferences of Alanine Dipeptide in Water, *J. Am. Chem. Soc.*, 2004, **126**, 2574–2581.
36 D. J. Tobias and C. L. Brooks, Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: a comparison of theoretical results, *J. Phys. Chem.*, 1992, **96**, 3864–3870.
37 S. Toal and R. Schweitzer-Stenner, Local order in the unfolded state: conformational biases and nearest neighbor interactions, *Biomolecules*, 2014, **4**, 725–773.
38 M. Kjaergaard, A. B. Norholm, R. Hendus-Altenburger, S. F. Pedersen, F. M. Poulsen and B. B. Kragelund, Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?, *Protein Sci.*, 2010, **19**, 1555–1564.
39 S. M. Kelly, T. J. Jess and N. Price, How to study proteins by circular dichroism, *Biochim. Biophys. Acta*, 2005, **1751**, 119–139.
40 J. Bandekar, Amide modes and protein conformation, *Biochim. Biophys. Acta*, 1992, **1120**, 123–143.
41 D. S. Wishart and B. D. Sykes, Chemical shifts as a tool for structure determination, *Methods Enzymol.*, 1994, **239**, 363–392.
42 P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, 1979.
43 K. W. Plaxco, C. J. Morton, S. B. Grimshaw, J. A. Jones, M. Pitkeathly, L. D. Campbell and C. M. Dobson, The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG, *J. Biomol. NMR*, 1997, **10**, 221–230.
44 A. S. Holehouse, K. Garai, N. Lyle, A. Vitalis and R. V. Pappu, Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation, *J. Am. Chem. Soc.*, 2015, **137**, 2984–2995.
45 V. N. Uversky, What does it mean to be natively unfolded?, *Eur. J. Biochem.*, 2002, **269**, 2–12.
46 A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine and R. V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 8183–8188.
47 S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels and B. Schuler, Charge interactions can dominate the dimensions of intrinsically disordered proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14609–14614.
48 M. R. Jensen, M. Zweckstetter, J. R. Huang and M. Blackledge, Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy, *Chem. Rev.*, 2014, **114**, 6632–6660.
49 M. R. Jensen, L. Salmon, G. Nodet and M. Blackledge, Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts, *J. Am. Chem. Soc.*, 2010, **132**, 1270–1272.
50 M. Schwalbe, V. Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M. R. Jensen, J. Biernat, S. Becker, E. Mandelkow, M. Zweckstetter and M. Blackledge, Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering, *Structure*, 2014, **22**, 238–249.
51 P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok and M. Blackledge, A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17002–17007.
52 Z. Shi, K. Chen, Z. Liu and N. R. Kallenbach, Conformation of the backbone in unfolded proteins, *Chem. Rev.*, 2006, **106**, 1877–1897.
53 R. Schweitzer-Stenner, Conformational propensities and residual structures in unfolded peptides and proteins, *Mol. BioSyst.*, 2012, **8**, 122–133.
54 H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels and B. Schuler, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 16155–16160.

55 Z. Shi, C. Anders Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, Polyproline II structure in a sequence of seven alanine residues, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9190–9195.

56 A. C. Wang and A. Bax, Determination of the backbone dihedral angles Φ in human ubiquitin from reparametrized empirical Karplus equations, *J. Am. Chem. Soc.*, 1996, **118**, 2483–2494.

57 J. S. Hu and A. Bax, Determination of φ and χ_1 angles in proteins from ^{13}C–^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond?, *J. Am. Chem. Soc.*, 1997, **119**, 6360–6368.

58 P. M. Cowan and S. Mc, Gavin, Structure of Poly-L-Proline, *Nature*, 1955, **176**, 501–503.

59 M. L. Tiffany and S. Krimm, New chain conformations of poly(glutamic acid) and polylysine, *Biopolymers*, 1968, **6**, 1379–1382.

60 W. L. Mattice, The effect of temperature and salt concentration on the circular dichroism exhibited by unionized derivatives of L-alanine in aqueous solution, *Biopolymers*, 1974, **13**, 169–183.

61 E. W. Ronish and S. Krimm, The calculated circular dichroism of polyproline ii in the polarizability approximation, *Biopolymers*, 1974, **13**, 1635–1651.

62 S. Woutersen and P. Hamm, Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy, *J. Phys. Chem. B*, 2000, **104**, 11316–11320.

63 J. Makowska, S. Rodziewicz-Motowidło, K. Bagińska, M. Makowski, J. A. Vila, A. Liwo, L. Chmurzyński and H. A. Scheraga, Further evidence for the absence of polyproline II stretch in the XAO peptide, *Biophys. J.*, 2007, **92**, 2904–2917.

64 J. Makowska, S. Rodziewicz-Motowidło, K. Bagińska, J. A. Vila, A. Liwo, L. Chmurzyński and H. A. Scheraga, Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1744–1749.

65 Z. Shi, K. Shen, Z. Liu and N. R. Kallenbach, Conformation in the Backbone in Unfolded Proteins, *Chem. Rev.*, 2006, **106**, 1877–1897.

66 J. Graf, P. H. Nguyen, G. Stock and H. Schwalbe, Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study, *J. Am. Chem. Soc.*, 2007, **129**, 1179–1189.

67 D. A. Case, C. Scheurer and R. Bruschweiler, Static and dynamic effects on vicinal scalar J couplings in proteins and peptides: a MD/DFT analysis, *J. Am. Chem. Soc.*, 2000, **122**, 10390–10397.

68 S. Zhang, R. Schweitzer-Stenner and B. Urbanc, Do Molecular Dynamics Force Fields Capture Conformational Dynamics of Alanine in Water?, *J. Chem. Theory Comput.*, 2020, **16**, 510–527.

69 Z. Shi, K. Chen, Z. Liu, A. Ng, W. C. Bracken and N. R. Kallenbach, Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β-sheet scales, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17964–17968.

70 D. M. Rogers, S. B. Jasim, N. T. Dyer, F. Auvray, M. Réfrégiers and J. D. Hirst, Electronic Circular Dichroism Spectroscopy of Proteins, *Chem.*, 2019, **5**, 2751–2774.

71 R. W. Woody, Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins, *Biopolymers*, 1978, **17**, 1451–1467.

72 F. Avbelj and R. L. Baldwin, Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5742–5747.

73 J. Grdadolnik, V. Mohacek-Grosev, R. L. Baldwin and F. Avbelj, Populations of the three major backbone conformations in 19 amino acid dipeptides, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 1794–1798.

74 A. G. Street and S. L. Mayo, Intrinsic β-sheet propensities result from van der Waals interactions between side chains and the local backbone, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9074–9076.

75 R. Schweitzer-Stenner, F. Eker, Q. Huang, K. Griebenow, P. A. Mroz and P. M. Kozlowski, Structure analysis of dipeptides in water by exploring and utilizing the structural sensitivity of amide III by polarized visible Raman, FTIR-spectroscopy and DFT based normal coordinate analysis, *J. Phys. Chem. B*, 2002, **106**, 4294–4304.

76 A. V. Mikhonin, Z. Ahmed, A. Ianoul and S. A. Asher, Assignments and conformational dependencies of the amide III peptide backbone UV resonance Raman bands, *J. Phys. Chem. B*, 2004, **108**, 19020–19028.

77 S. A. Asher, A. Ianoul, G. Mix, M. N. Boyden, A. Karnoup, M. Diem and R. Schweitzer-Stenner, Dihedral ψ angle dependence of the amide III vibration: a uniquely sensitive UV resonance Raman secondary structural probe, *J. Am. Chem. Soc.*, 2001, **123**, 11775–11781.

78 S. Krimm and J. Bandekar, Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins, *Adv. Protein Chem.*, 1986, **38**, 181–364.

79 R. Schweitzer-Stenner, Secondary structure analysis of polypeptides based on an excitonic coupling model to describe the band profile of amide I′ of IR, Raman, and vibrational circular dichroism spectra, *J. Phys. Chem. B*, 2004, **108**, 16965–16975.

80 C. M. Baronio and A. Barth, The Amide i Spectrum of Proteins – Optimization of Transition Dipole Coupling Parameters Using Density Functional Theory Calculations, *J. Phys. Chem. B*, 2020, **124**, 1703–1714.

81 R. Schweitzer-Stenner, Distribution of conformations sampled by the central amino acid residue in tripeptides inferred from amide i band profiles and NMR scalar coupling constants, *J. Phys. Chem. B*, 2009, **113**, 2922–2932.

82 F. Eker, X. Cao, L. Nafie and R. Schweitzer-Stenner, Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR, and VCD spectroscopy study, *J. Am. Chem. Soc.*, 2002, **124**, 14330–14341.

83 L. He, A. E. Navarro, Z. Shi and N. R. Kallenbach, End effects influence short model peptide conformation, *J. Am. Chem. Soc.*, 2012, **134**, 1571–1576.

84 O. Lee, G. M. Roberts and M. Diem, IR vibrational CD in alanyl tripeptide: indication of a stable solution conformer, *Biopolymers*, 1989, **28**, 1759–1770.
85 S. Toal, D. Meral, D. Verbaro, B. Urbanc and R. Schweitzer-Stenner, PH-independence of trialanine and the effects of termini blocking in short peptides: a combined vibrational, NMR, UVCD, and molecular dynamics study, *J. Phys. Chem. B*, 2013, **117**, 3689–3706.
86 S. Gnanakaran and A. E. Garcia, Validation of an All-Atom Protein Force Field: From Dipeptides to Larger Peptides, *J. Phys. Chem. B*, 2003, **107**, 12555–12557.
87 A. E. Garcia, Characterization of non-alpha helical conformations in Ala peptides, *Polymer*, 2004, **45**, 669–676.
88 C. D. Poon, E. T. Samulski, C. F. Weise and J. C. Weisshaar, Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution?, *J. Am. Chem. Soc.*, 2000, **122**, 5642–5643.
89 A. Kentsis, M. Mezei, T. Gindin and R. Osman, Unfolded State of Polyalanine Is a Segmented Polyproline II Helix, *Proteins: Struct., Funct., Genet.*, 2004, **55**, 493–501.
90 A. Hagarman, D. Mathieu, S. Toal, T. J. Measey, H. Schwalbe and R. Schweitzer-Stenner, Amino acids with hydrogen-bonding side chains have an intrinsic tendency to sample various turn conformations in aqueous solution, *Chem. – Eur. J.*, 2011, **17**, 6789–6797.
91 A. Hagarman, T. J. Measey, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I' band profiles and NMR scalar coupling constants, *J. Am. Chem. Soc.*, 2010, **132**, 540–551.
92 R. Schweitzer-Stenner, A. Hagarman, S. Toal, D. Mathieu and H. Schwalbe, Disorder and order in unfolded and disordered peptides and proteins: a view derived from tripeptide conformational analysis. I. Tripeptides with long and predominantly hydrophobic side chains, *Proteins Struct., Funct., Bioinf.*, 2013, **81**, 955–967.
93 K. Rybka, S. E. Toal, D. J. Verbaro, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, Disorder and order in unfolded and disordered peptides and proteins: a view derived from tripeptide conformational analysis. II. Tripeptides with short side chains populating asx and β-type like turn conformations, *Proteins Struct., Funct., Bioinf.*, 2013, **81**, 968–983.
94 V. Muñoz and L. Serrano, Intrinsic secondary structure propensities of the amino acids, using statistical ϕ–ψ matrices: comparison with experimental scales, *Proteins Struct., Funct., Bioinf.*, 1994, **20**, 301–311.
95 M. B. Swindells, M. W. Macarthur and J. M. Thornton, Intrinsic φ, ψ propensities of amino acids, derived from the coil regions of known structures, *Nat. Struct. Biol.*, 1995, **2**, 569–603.
96 A. K. Jha, A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick and K. F. Freed, Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library, *Biochemistry*, 2005, **44**, 9691–9702.
97 M. C. Baxa, E. J. Haddadian, A. K. Jha, K. F. Freed and T. R. Sosnick, Context and force field dependence of the loss of protein backbone entropy upon folding using realistic denatured and native state ensembles, *J. Am. Chem. Soc.*, 2012, **134**, 15929–15936.
98 D. Ting, G. Wang, M. Shapovalov, R. Mitra, M. I. Jordan and R. L. Dunbrack, Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model, *PLoS Comput. Biol.*, 2010, **6**, e1000763.
99 L. L. Perskie, T. O. Street and G. D. Rose, Structures, basins, and energies: a deconstruction of the Protein Coil Library, *Protein Sci.*, 2008, **17**, 1151–1161.
100 T. R. Sosnick, Sampling library, **http://godzilla.uchicago.edu/cgi-bin/rama.cgi**.
101 B. Milorey, H. Schwalbe, N. O'Neill and R. Schweitzer-Stenner, Repeating Aspartic Acid Residues Prefer Turn-like Conformations in the Unfolded State: Implications for Early Protein Folding, *J. Phys. Chem. B*, 2021, **125**, 11392–11407.
102 R. Schweitzer-Stenner and S. E. Toal, Entropy reduction in unfolded peptides (and proteins) due to conformational preferences of amino acid residues, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22527–22536.
103 S. Tanaka and H. A. Scheraga, Statistical Mechanical Treatment of Protein Conformation. 4. A Four-State Model for Specific-Sequence Copolymers of Amino Acids, *Macromolecules*, 1976, **9**, 812–833.
104 A. L. Rucker and T. P. Creamer, Polyproline II helical structure in protein unfolded states: lysine peptides revisited, *Protein Sci.*, 2002, **11**, 980–985.
105 A. L. Rucker, C. T. Pager, M. N. Campbell, J. E. Qualls and T. P. Creamer, Host-guest scale of left-handed polyproline II helix formation, *Proteins: Struct., Funct., Genet.*, 2003, **53**, 68–75.
106 S. J. Whittington, B. W. Chellgren, V. M. Hermann and T. P. Creamer, Urea promotes polyproline II helix formation: implications for protein denatured states, *Biochemistry*, 2005, **44**, 6269–6275.
107 A. V. Mikhonin, N. S. Myshakina, S. V. Bykov and S. A. Asher, UV resonance Raman determination of polyproline II, extended 2.5 1-helix, and β-sheet ψ angle energy landscape in poly-L-lysine and poly-L-glutamic acid, *J. Am. Chem. Soc.*, 2005, **127**, 7712–7720.
108 Z. P. Gates, M. C. Baxa, W. Yu, J. A. Riback, H. Li, B. Roux, S. B. H. Kent and T. R. Sosnick, Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 2241–2246.
109 J. Danielsson, J. Jarvet, P. Damberg and A. Gräslund, The Alzheimer β-peptide shows temperature-dependent transitions between left-handed 31-helix, β-strand and random coil secondary structure, *FEBS J.*, 2005, **272**, 3938–3949.
110 R. Schweitzer-Stenner and S. E. Toal, Anticooperative Nearest-Neighbor Interactions between Residues in Unfolded Peptides and Proteins, *Biophys. J.*, 2018, **114**, 1046–1057.
111 A. K. Jha, A. Colubri, K. F. Freed and T. R. Sosnick, Statistical coil model of the unfolded state: resolving the

reconciliation problem, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13099–13104.
112 H. T. Tran, X. Wang and R. V. Pappu, Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins, *Biochemistry*, 2005, **44**, 11369–11380.
113 N. C. Fitzkee and G. D. Rose, Reassessing random-coil statistics in unfolded proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12497–12502.
114 B. Zagrovic, J. Lipfert, E. J. Sorin, I. S. Millett, W. F. Van Gunsteren, S. Doniach and V. S. Pande, Unusual compactness of a polyproline type II structure, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11698–11703.
115 W. G. Han, K. J. Jalkanen, M. Elstner and S. Suhai, Theoretical study of aqueous *N*-acetyl-L-alanine *N*′-methylamide: structures and Raman, VCD, and ROA spectra, *J. Phys. Chem. B*, 1998, **102**, 2587–2602.
116 S. Gnanakaran and A. E. García, Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures, *Proteins: Struct., Funct., Genet.*, 2005, **59**, 773–782.
117 M. Mezei, P. J. Fleming, R. Srinivasan and G. D. Rose, Polyproline II Helix Is the Preferred Conformation for Unfolded Polyalanine in Water, *Proteins: Struct., Funct., Genet.*, 2004, **55**, 502–507.
118 Z. Shi, R. W. Woody and N. R. Kallenbach, Is polyproline II a major backbone conformation in unfolded proteins?, *Adv. Protein Chem.*, 2002, **62**, 163–240.
119 S. E. Toal, D. J. Verbaro and R. Schweitzer-Stenner, Role of enthalpy-entropy compensation interactions in determining the conformational propensities of amino acid residues in unfolded peptides, *J. Phys. Chem. B*, 2014, **118**, 1309–1318.
120 F. Avbelj and R. L. Baldwin, Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5742–5747.
121 G. Lanza and M. A. Chiacchio, Effects of hydration on the zwitterion trialanine conformation by electronic structure theory, *J. Phys. Chem. B*, 2016, **120**, 11705–11719.
122 G. Lanza and M. A. Chiacchio, Comprehensive and accurate ab initio energy surface of simple alanine peptides, *ChemPhysChem*, 2013, **14**, 3284–3293.
123 G. Lanza and M. A. Chiacchio, Interfacial water at the trialanine hydrophilic surface: a DFT electronic structure and bottom-up investigation, *Phys. Chem. Chem. Phys.*, 2015, **17**, 17101–17111.
124 N. V. Ilawe, A. E. Raeber, R. Schweitzer-Stenner, S. E. Toal and B. M. Wong, Assessing backbone solvation effects in the conformational propensities of amino acid residues in unfolded peptides, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24917–24924.
125 D. Meral, S. Toal, R. Schweitzer-Stenner and B. Urbanc, Water-Centered Interpretation of Intrinsic pPII Propensities of Amino Acid Residues: In Vitro-Driven Molecular Dynamics Study, *J. Phys. Chem. B*, 2015, **119**, 13237–13251.
126 B. Andrews, J. Guerra, R. Schweitzer-Stenner and B. Urbanc, Do molecular dynamics force fields accurately model Ramachandran distributions of amino acid residues in water?, *Phys. Chem. Chem. Phys.*, 2022, **24**, 3259–3279.
127 F. Eker, K. Griebenow and R. Schweitzer-Stenner, Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy, *J. Am. Chem. Soc.*, 2003, **125**, 8178–8185.
128 Z. Liu, K. Chen, A. Ng, Z. Shi, R. W. Woody and N. R. Kallenbach, Solvent dependence of PII conformation in model alanine peptides, *J. Am. Chem. Soc.*, 2004, **126**, 15141–15150.
129 B. Milorey, S. Farrell, S. E. Toal and R. Schweitzer-Stenner, Demixing of water and ethanol causes conformational redistribution and gelation of the cationic GAG tripeptide, *Chem. Commun.*, 2015, **51**, 16498–16501.
130 D. Diguiseppi, B. Milorey, G. Lewis, N. Kubatova, S. Farrell, H. Schwalbe and R. Schweitzer-Stenner, Probing the Conformation-Dependent Preferential Binding of Ethanol to Cationic Glycylalanylglycine in Water/Ethanol by Vibrational and NMR Spectroscopy, *J. Phys. Chem. B*, 2017, **121**, 5744–5758.
131 B. Andrews, S. Zhang, R. Schweitzer-Stenner and B. Urbanc, Glycine in Water Favors the Polyproline II State, *Biomolecules*, 2020, **10**, 1121.
132 R. Schweitzer-Stenner, Exploring Nearest Neighbor Interactions and Their Influence on the Gibbs Energy Landscape of Unfolded Proteins and Peptides, *Int. J. Mol. Sci.*, 2022, **23**, 5643.
133 K. A. Dill, Additivity principles in biochemistry, *J. Biol. Chem.*, 1997, **272**, 701–704.
134 C. J. Penkett, C. Redfield, I. Dodd, J. Hubbard, D. L. McBay, D. E. Mossakowska, R. A. G. Smith, C. M. Dobson and L. J. Smith, NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein, *J. Mol. Biol.*, 1997, **274**, 152–159.
135 F. Avbelj and R. L. Baldwin, Origin of the neighboring residue effect on peptide backbone conformation, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10967–10972.
136 M. H. Zaman, M. Y. Shen, R. S. Berry, K. F. Freed and T. R. Sosnick, Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the flory isolated-pair hypothesis for peptides, *J. Mol. Biol.*, 2003, **331**, 693–711.
137 J. Debartolo, A. Jha, K. F. Freed and T. R. Sosnick, in *Protein and Peptide Folding, Misfolding, and Non-Folding*, ed. R. Schweitzer-Stenner, Wiley & Sons, Chichester, 2012, pp. 79–98.
138 S. E. Toal, N. Kubatova, C. Richter, V. Linhard, H. Schwalbe and R. Schweitzer-Stenner, Randomizing the Unfolded State of Peptides (and Proteins) by Nearest Neighbor Interactions between Unlike Residues, *Chem. – Eur. J.*, 2015, **21**, 5173–5192.
139 B. Milorey, R. Schweitzer-Stenner, B. Andrews, H. Schwalbe and B. Urbanc, Short peptides as predictors for the

structure of polyarginine sequences in disordered proteins, *Biophys. J.*, 2021, **120**, 662–676.
140 R. Schweitzer-Stenner, B. Milorey and H. Schwalbe, Randomizing of Oligopeptide Conformations by Nearest Neighbor Interactions between Amino Acid Residues, *Biomolecules*, 2022, **12**, 684.
141 S. E. Toal, N. Kubatova, C. Richter, V. Linhard, H. Schwalbe and R. Schweitzer-Stenner, Corrigendum: Randomizing the Unfolded State of Peptides (and Proteins) by Nearest Neighbor Interactions between Unlike Residues, *Chemistry*, 2017, **23**, 18084–18087.
142 R. Schweitzer-Stenner and S. E. Toal, Construction and comparison of the statistical coil states of unfolded and intrinsically disordered proteins from nearest-neighbor corrected conformational propensities of short peptides, *Mol. BioSyst.*, 2016, **12**, 3294–3306.
143 F. Avbelj, S. G. Grdadolnik, J. Grdadolnik and R. L. Baldwin, Intrinsic backbone preferences are fully present in blocked amino acids, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1272–1277.
144 R. B. Best and G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, *J. Phys. Chem. B*, 2009, **113**, 9004–9015.
145 R. B. Best, N. V. Buchete and G. Hummer, Are current molecular dynamics force fields too helical?, *Biophys. J.*, 2008, **95**, L07–L09.
146 D. Verbaro, I. Ghosh, W. M. Nau and R. Schweitzer-Stenner, Discrepancies between conformational distributions of a polyalanine peptide in solution obtained from molecular dynamics force fields and amide I′ band profiles, *J. Phys. Chem. B*, 2010, **114**, 17201–17208.
147 P. S. Nerenberg and T. Head-Gordon, Optimizing protein–solvent force fields to reproduce intrinsic conformational preferences of model peptides, *J. Chem. Theory Comput.*, 2011, **7**, 1220–1230.
148 S. C. Lovell, J. M. Word, J. S. Richardson and D. C. Richardson, The penultimate rotamer library, *Proteins: Struct., Funct., Genet.*, 2000, **40**, 389–408.
149 C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu and C. Simmerling, Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, *J. Chem. Theory Comput.*, 2020, **16**, 528–552.
150 A. Vitalis and R. V. Pappu, ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions, *J. Comput. Chem.*, 2009, **30**, 673–699.
151 S. Li, C. T. Andrews, T. Frembgen-Kesner, M. S. Miller, S. L. Siemonsma, T. D. Collingsworth, I. T. Rockafellow, N. A. Ngo, B. A. Campbell, R. F. Brown, C. Guo, M. Schrodt, Y. T. Liu and A. H. Elcock, Molecular dynamics simulations of 441 two-residue peptides in aqueous solution: conformational preferences and neighboring residue effects with the amber ff99sb-ildn-NMR force field, *J. Chem. Theory Comput.*, 2015, **11**, 1315–1329.
152 S. Li and A. H. Elcock, Residue-Specific Force Field (RSFF2) improves the modeling of conformational behavior of peptides and proteins, *J. Phys. Chem. Lett.*, 2015, **6**, 2127–2133.
153 Y. S. Jung, K. I. Oh, G. S. Hwang and M. Cho, Neighboring residue effects in terminally blocked dipeptides: implications for residual secondary structures in intrinsically unfolded/disordered proteins, *Chirality*, 2014, **26**, 443–452.
154 Y. Yuan and F. Wang, Dipole Cooperativity and Polarization Frustration Determine the Secondary Structure Distribution of Short Alanine Peptides in Water, *J. Phys. Chem. B*, 2023, **127**(14), 3126–3138.
155 F. Jiang, W. Han and Y. D. Wu, Influence of side chain conformations on local conformational features of amino acids and implication for force field development, *J. Phys. Chem. B*, 2010, **114**, 5840–5850.
156 C. Xu, J. Wang and H. Liu, A Hamiltonian replica exchange approach and its application to the study of side-chain type and neighbor effects on peptide backbone conformations, *J. Chem. Theory Comput.*, 2008, **4**, 1348–1359.
157 D. A. C. Beck, D. O. V. Alonso, D. Inoyama and V. Daggett, The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 12259–12264.
158 P. B. Law and V. Daggett, The relationship between water bridges and the polyproline II conformation: a large-scale analysis of molecular dynamics simulations and crystal structures, *Protein Eng., Des. Sel.*, 2010, **23**, 27–33.
159 D. Braun, G. Wider and K. Wüthrich, Sequence-Corrected ^{15}N 'Random Coil' Chemical Shifts, *J. Am. Chem. Soc.*, 1994, **116**, 8466–8469.
160 G. Merutka, H. Jane Dyson and P. E. Wright, 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG, *J. Biomol. NMR*, 1995, **5**, 14–24.
161 A. Bundi and K. Wüthrich, ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H–Gly–Gly–X–L–Ala–OH, *Biopolymers*, 1979, **18**, 285–297.
162 F. Avbelj, in *Protein and Peptide Folding, Misfolding, and Non-Folding*, ed. R. Schweitzer-Stenner, John Wiley & Sons, Hobocken, 2012, pp. 131–158.
163 F. Avbelj, D. Kocjan and R. L. Baldwin, Protein chemical shifts arising from α-helices and β-sheets depend on solvent exposure, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17394–17397.
164 D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR*, 1995, **5**, 67–81.
165 S. Schwarzinger, G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright and H. J. Dyson, Sequence-dependent correction of random coil NMR chemical shifts, *J. Am. Chem. Soc.*, 2001, **123**, 2970–2978.
166 M. Kjaergaard and F. M. Poulsen, Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution, *J. Biomol. NMR*, 2011, **50**, 157–165.

167 K. Tamiola, B. Acar and F. A. A. Mulder, Sequence-specific random coil chemical shifts of intrinsically disordered proteins, *J. Am. Chem. Soc.*, 2010, **132**, 18000–18003.

168 H. J. Dyson and P. E. Wright, Insights into the Structure and Dynamics of Unfolded Proteins from Nuclear Magnetic Resonance, *Adv. Protein Chem.*, 2002, **62**, 311–340.

169 B. H. Zimm and J. K. Bragg, Theory of the phase transition between helix and random coil in polypeptide chains, *J. Chem. Phys.*, 1959, **31**, 526–535.

170 S. Lifson and A. Roig, On the theory of helix-coil transitions in polypeptides, *J. Chem. Phys.*, 1961, **34**, 1963–1974.

171 T. P. Creamer and G. D. Rose, α-Helix-forming propensities in peptides and proteins, *Proteins Struct., Funct., Bioinf.*, 1994, **19**, 85–97.

172 A. Chakrabartty, T. Kortemme and R. L. Baldwin, Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions, *Protein Sci.*, 1994, **3**, 843–852.

173 B. Basel, Side Chain Conformations, **https://swissmodel.expasy.org/course/text/chapter3.htm**.

174 C.-L. Towse, S. J. Rysavy, I. M. Vulovic and V. Daggett, New Dynamics Rotamer Libraries: Data Driven Analysis of Side Chain Conformational Propensities, *Structure*, 2016, **24**, 187–199.

175 J. C. Mortensen, J. Damjanovic, J. Miao, T. Hui and Y. S. Lin, A backbone-dependent rotamer library with high (ϕ, ψ) coverage using metadynamics simulations, *Protein Sci.*, 2022, **31**, e4491.

176 M. V. Shapovalov and R. L. Jr Dunbrack, A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions, *Structure*, 2011, **19**, 844–858.

177 P. Chakrabarti and D. Pal, The interrelationships of side-chain and main-chain conformations in proteins, *Prog. Biophys. Mol. Biol.*, 2001, **76**, 1–102.

178 L. Duitch, S. Toal, T. J. Measey and R. Schweitzer-Stenner, Triaspartate: a model system for conformationally flexible DDD motifs in proteins, *J. Phys. Chem. B*, 2012, **116**, 5160–5171.

179 M. C. Baxa, E. J. Haddadian, J. M. Jumper, K. F. Freed and T. R. Sosnick, Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 15396–15401.

180 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

181 C. A. Kim and J. M. Berg, Thermodynamic β-sheet propensities measured using a zinc-finger host peptide, *Nature*, 1993, **362**, 267–270.