Impact of the Conformational Variability of Oligopeptides on the Computational Prediction of their CD Spectra

M. Michaelis,[*,†,‡] N. Hildebrand,[†] R. H. Meißner,[†] N. Wurzler,[†] Z. Li,[¶] J. D. Hirst,[¶] A. Micsonai,[§] J. Kardos,[§] M. Delle Piane,[*,†] and L. Colombi Ciacchi[†]

†*Hybrid Materials Interfaces Group, University of Bremen, Faculty of Production Engineering, Bremen Center for Computational Materials Science, Center for Environmental Research and Sustainable Technology (UFT), and MAPEX Center for Materials and Processes, Am Fallturm 1, 28359 Bremen, Germany*

‡*Biomolecular and Materials Interface Research Group, Interdisciplinary Biomedical Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom*

¶*School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

§*Department of Biochemistry, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest, H-1117 Hungary*

E-mail: michaelis@uni-bremen.de; massimo.dellepiane@hmi.uni-bremen.de

Abstract

While successful in the structural determination of ordered biomolecules, the spectroscopic investigation of oligopeptides in solution is hindered by their complex and rapidly changing conformational ensemble. The measured circular dichroism (CD) spectrum of an oligopeptide is an average of the signals coming from all the ensemble of microstates, severely limiting its interpretation, in contrast to the successful structural determination of ordered biomolecules. Spectral deconvolution methods to estimate the secondary structure contributions in the ensemble are still mostly based on databases of larger ordered proteins. Here we establish how the interpretation of CD spectra of oligopeptides can be enhanced by the ability to compute the same observable from a set of atomic coordinates. Focusing on two representative oligopeptides featuring a known propensity towards an α-helical and β-hairpin motif, respectively, we compare and cross-validate the structural information coming from: deconvolution of the experimental CD spectra, sequence-based *de novo* structure prediction and molecular dynamics simulations based on enhanced sampling methods. We find that small conformational variations can give rise to significant changes in the CD signals. Therefore, while for the simpler conformational landscape of the α-helical peptide *de novo* structure prediction can already give reasonable agreement with the experiment, an extended ensemble of conformers needs to be considered for the β-hairpin sequence.

Keywords

Oligopeptide, Circular dichroism, Molecular dynamics, Free energy, Conformational ensemble

Introduction

Oligopeptides and disordered proteins are characterized by an ensemble distribution of dynamic and rapidly changing conformers (microstates).[1] Fundamental understanding of such

conformational ensembles is key to the basic principles of protein folding and misfolding, with important implications for the aggregation of proteins and their related diseases.[2,3] In these cases, solid-state based techniques such as X-ray diffraction, while extremely successful in the structural determination of ordered proteins, are not informative, because they miss a full consideration of the ensemble. Solution-based experimental structural investigation methods, e.g. NMR or optical spectroscopies, also deliver incomplete information, since they output a signal representative of the conformational macrostate, *i.e.* the Boltzmann average of all ensemble microstates. However, the measured signals can be commonly analysed to estimate of the relative population of typical secondary structure elements within this ensemble using deconvolution approaches.[4–7]

Among other solution-based methods, electronic circular cichroism (CD) spectroscopy is widely used for identifying and quantifying the secondary structures of solvated biomolecules, using regression techniques based on datasets of proteins with both accessible CD spectra and crystal structures.[8] A more direct observation of the conformations of oligopeptides can be achieved *via* computational approaches such as all-atom molecular dynamics (MD) based on classical force fields, given their proven ability to predict the dynamics of biomolecular systems with high accuracy. However, so-called enhanced-sampling MD methods are required to capture with statistical relevance all rare events contributing to intramolecular conformational changes, and thus to explore the conformational ensemble of oligopetides.[9] In particular, Replica Exchange with Solute Tempering (REST)[10] in combination with Metadynamics (MetaD)[11] is an advanced-sampling MD method that has been successfully used for this purpose.[12–15]

In this context, the ability to compute experimental observables from RESTMetaD trajectories constitutes a valuable tool of cross-validation and can strengthen the interpretation of the experimental data. The combination of CD spectroscopy and atomistic simulations has been indeed successfully employed to gain some atomistic understanding of conformational ensembles.[14,15] This is due to the fact that a CD spectrum can be computationally predicted

based on the atomic positions of a given biomolecular conformation.[16–18] When dealing with ensembles, however, the need to consider thousands of different conformations requires a computationally efficient approach for predicting the CD spectra of the corresponding microstates. A popular approach is provided by the DichroCalc software platform,[19] which employs the so-called matrix method to calculate CD ellipticities from first principles using exciton theory.

While promising in general, this approach has been proven to work best for oligopeptides and proteins with high α-helical secondary structure content.[14,15] This is due to the nature of the electronic transitions in CD, which can be well captured by only a few matrix elements in the case of α-helices.[19] Moreover, the regularity of the α-helical hydrogen-bond pattern produces intense spectral features in well-defined wavelength regions, facilitating the computational prediction.[6] On the quest towards a generalization of this approach, it would be desirable to tackle the higher structural variability of conformations associated with other secondary structure motifs, in particular β-sheets and turns.[4,5]

The aim of this paper is to move in this direction, by choosing two oligopetides featuring, respectively, a preferential α-helical and a β-hairpin conformation, and determining both experimentally and theoretically their CD spectra in order to evaluate and compare their conformational ensembles. For the helical model peptide, we choose the 4DAR5 sequence,[20,21] which some of us investigated previously,[14] demonstrating for the first time the possibility of predicting experimental CD spectra from enhanced-sampling MD simulations. For the β-hairpin model, we choose the C-terminal fragment of the B1 domain of protein G (called GB1 throughout this paper), which has been intensely studied experimentally and theoretically.[22–30] Our analysis will focus both on rapid *de novo* secondary structure prediction tools and more demanding RESTMetaD simulations, from which CD spectra will be calculated and compared with our own experiments.

In doing so, we will attempt to separately assess the predictive power and limitations of two different key elements involved in the theoretical CD prediction, namely (i) the pa-

rameter sets of the DichroCalc method, and (ii) the influene of the force fields used in the MD simulations. The latter is crucial to predict the correct equilibrium distribution of conformational microstates, a task that has proven very hard especially for short peptides with non-helical structures. The former should lead to quantitatively correct spectra when applied to the correct ensemble of microstates. The consideration of peptides with known conformational phase spaces will allow us to distinguish between these two potential sources of errors and thus help us apply the same methodology to new oligopeptide sequences with unknown conformations in future studies.

Materials and Methods

Circular dichroism (CD) experiments

CD experiments were performed with oligopeptides with sequences Ac-DDDDAAAAARRRR-Am (4DAR5), GEWTYDDATKTFTVTE (GB1-uncapped) and Ac-GEWTYDDATKTFTVTE-Am (GB1) synthezised by JPT (purity $\geq 95\,\%$ Berlin, Germany). The as-purchased oligopeptide 4DAR5 was dissolved in ddH_2O ($18\,\mathrm{M\Omega\,cm^{-1}}$, MilliQ, Synergy, Millipore, Germany) at a stock concentration of $1\,\mathrm{mg\ ml^{-1}}$, as quantified by absorbance at 205 and 214 nm.[31,32] The as-purchased oligopeptides GB1 and GB1-capped were dissolved in a 3 mM PBS solution (41 mM NaCl; P4417, Sigma Aldrich, Germany) at a concentration of $1\,\mathrm{mg\ ml^{-1}}$, as quantified by absorbance at 205 and 214 nm.

CD spectra were recorded with an Applied Photophysics Chirascan spectrometer running the Pro-Data Chirascan software (v4.2.22). At least three repeat scans for each sample were measured at 25°C, over the wavelength range of 190 to 250 nm using intervals of 1 nm, in Suprasil quartz cells (Hellma UK Ltd.) with a pathlength of 0.05 cm. The scans were averaged, the respective baseline subtracted and the resulting net spectra smoothed with a Savitsky-Golay Filter using smoothing windows of 5 to 10 data points.

The mean residue ellipticity (Θ_{MRE}) was defined as[33]

$$\Theta_{MRE} = \frac{\Theta}{c \cdot l \cdot n} \tag{1}$$

where Θ is the raw CD ellipticity (mdeg), n is the number of amino acids in the solvated peptide, l is the pathlength of the used quartz cuvette (mm) and c the molar concentration of the peptides. To estimate the relative amount of specific secondary conformational elements in the samples, the CD spectra were analysed using the BeStSel webserver.[4,5] This analysis decomposes the whole CD spectrum in terms of individual secondary structure contributions using an extended set of eight independent elements with special emphasis on the different types of β-sheets defined by different twist angles in the polypeptide backbone.[4,5]

De novo secondary structure prediction

The intrinsic conformational propensities of the two investigated peptides are evaluated with the PEP-Fold secondary structure prediction server.[34–36] PEP-Fold first predicts Structural Alphabet (SA) letters from the primary peptide sequence using a hidden Markov model approach. The letter fragments are then assembled by a Greedy procedure driven by a modified OPEP coarse-grained force field energy score followed by a clustering procedure, leading to a maximum of five most-probable 3D structures.

Enhanced sampling Molecular Dynamics (MD) simulations

MD simulations were performed with LAMMPS,[37] employing the Amber03 force field,[38,39] following Meissner et al.[14] In the GB1 case, a further simulation employing the Amber14SB force field was performed, for comparison.[40] Visualization and analysis of the trajectories were performed with VMD.[41] Free energy profiles for both peptides, in capped form, along the respectively chosen collective variables were computed with Replica Exchange with Solute Tempering (REST)[10,42,43] combined with Metadynamics (MetaD),[11] as employed in

previous works.[13,14] Both peptides were defined as the 'solute' in the REST simulations, whose temperature was scaled in the different system replicas ('hot' system region). The water and the salt ions remained at the base temperature $T_0 = 300\,\mathrm{K}$ ('cold' system region). For the peptide 4DAR5 we used seven replicas at temperatures T_i corresponding to 300, 325, 350, 375, 400, 425 and 450 K, respectively.[14] For the peptide GB1 ten replicas were applied using temperatures of 300, 328, 360, 394, 432, 474, 519, 569, 624 and 684 K following a geometric distribution. Defining $\beta_i = 1/(k_\mathrm{B}T_i)$, where k_B is the Boltzmann constant, the Lennard-Jones parameters ϵ of the hot atoms in the i-th replica were scaled by the factor β_i/β_0, and their charges q by the factor $\sqrt{\beta_i/\beta_0}$. Of the bonded interactions, only the dihedral force constants were scaled. Exchanges between the replicas were attempted every 0.5 ps, following a Metropolis-Hastings acceptance criterion.

Further conformational sampling was achieved by means of the 'well-tempered' metadynamics method introduced by Barducci et al.,[11] with a bias factor of 10. For the peptide 4DAR5 the helicity H was used as the collective variable (CV).[14] This was defined, as in PLUMED 1.3,[44] by the deviation of the dihedral angles Φ_j and Ψ_j from the ideal angles of an α-helix, $\Phi_0 = -68.75°$ and $\Psi_0 = -45.0°$:[14]

$$H = \frac{1}{N-2}\sum_{i=2}^{N-1}\prod_{j=i-1}^{i+1}\frac{1}{4}\big[\cos(\Phi_j - \Phi_0) + 1\big]\big[\cos(\Psi_j - \Psi_0) + 1\big] \tag{2}$$

with N corresponding to the total number of amino acids in the helix ($N = 13$) and the index i spanning between 11 residues. An ideal helix assumes a value $H = 100\,\%$, whereas a completely unordered structure assumes a value $H = 0\,\%$. In accordance with Bussi et al.,[12] for GB1 the radius of gyration R_{Gyr} and the number of hydrogen bonds N_H were used as the CVs. They are defined in PLUMED 1[44] as follows:

$$R_{Gyr} = \left(\frac{\sum_i^n m_i |r_i - r_{COM}|^2}{\sum_i^n m_i}\right)^{\frac{1}{2}} \quad , \tag{3}$$

where m_i and m_i are the positions and masses of the heavy atoms of the backbone of the

peptide, respectively, and r_{COM} is the position of the center of mass;

$$N_H = \sum_{i \in O} \sum_{j \in H} \frac{1 - \left(\frac{(r_i - r_j)}{d_0}\right)^6}{1 - \left(\frac{(r_i - r_j)}{d_0}\right)^{12}} \quad , \tag{4}$$

with $d_0 = 2.5$ Å, and the sum running only over the O and H atoms of the peptide backbone.[12] Along all CVs, Gaussian hills with a height of of 0.7 kcal mol^{-1} were deposited with a frequency of 0.5 ps^{-1}. The widths of the Gaussian hills were 0.1, 0.25 and 0.05 Å for H, R_{Gyr} and N_H, respectively.

Structural and cluster analysis

Cluster analysis of the conformational microstates sampled by the RESTmetaD trajectories in selected regions of the free energy landscapes was performed using the GROMOS algorithm,[45] as implemented in Gromacs, according to the differences in the root-mean square displacement (RMSD) values of individual conformers, using an RMSD cutoff of 3 Å applied to the atoms of the backbone. The conformers were aligned by a rigid rotation and translation to minimize the RMSD variation prior to clustering. For comparison, a hierarchical clustering was performed, on selected regions, using the distance in inter-residue contact maps. For further validation, average inter-residue contact maps were also computed for the most populated clusters obtained by backbone RMSD clustering. The secondary structures of the most populated clusters were also analyzed *via* the STRIDE algorithm.[46] The Φ and Ψ dihedrals of the peptides backbones along the trajectories were computed *via* the 'chi' program in Gromacs to create Ramachandran plots.

Calculation of theoretical CD spectra and analysis

Theoretical CD spectra of individual conformational microstates were calculated using the DichroCalc web interface.[17,19] DichroCalc estimates the mutual interactions between electronic transitions with a model exciton Hamiltonian matrix;[19] the sign and magnitude of

the interactions is dictated by the relative orientation and separation of chromophores, *i.e.*, by the precise structure of the protein/peptide. The matrix elements were computed using three different parameter sets modelling the $n \rightarrow \pi^*$ and $\pi_{nb} \rightarrow \pi^*$ transitions of the peptide bond chromophore, namely: a set derived from semi-empirical calculations,[47] a set based on *ab initio* calculations,[48] and another *ab initio* set in which the vibrational structure of the $\pi_{nb} \rightarrow \pi^*$ transition is explicitly included.[49] The vibrational structure, arising from the Franck Condon progression of the $\pi_{nb} \rightarrow \pi^*$ transition, contributes to the broadening of CD bands and is thus important to improve the agreement with experimental spectra. In all calculations, neither the $\pi_b \rightarrow \pi^*$ transition nor charge-transfer transitions were included. The calculated line spectra were convoluted with Gaussian bands of bandwidth 12.5 nm, except for the spectra calculated with the parameters incorporating vibrational structure, where a narrower bandwidth of 10 nm was applied, to avoid double-counting.

Results

Circular dichroism experiments

We first report the experimentally measured CD spectra of the two model peptides 4DAR5 and GB1. In agreement with previous investigations,[20,21] the CD spectrum of 4DAR5 is characterized by two minima at 208 and 222 nm and a maximum at 190 nm, an overall shape which is commonly associated with helical conformations (Figure 1 a). However, the BestSel analysis of the spectrum reveals a helical content of only 24.5 ± 0.8 % and a major contribution of unordered conformations ('other' elements). The CD spectrum of GB1 is reported in Figure 1 b. Since a capping of the end groups has been reported to stabilize the β-hairpin conformation of the peptide,[50] we have also measured the CD signal for the native form of GB1 (GB1-uncapped, Figure S1 in Supporting Information). However, we do not report any difference between the two forms, suggesting that the acetyl-amidation does not have a significant effect on the conformational ensemble. In the following, we will only refer

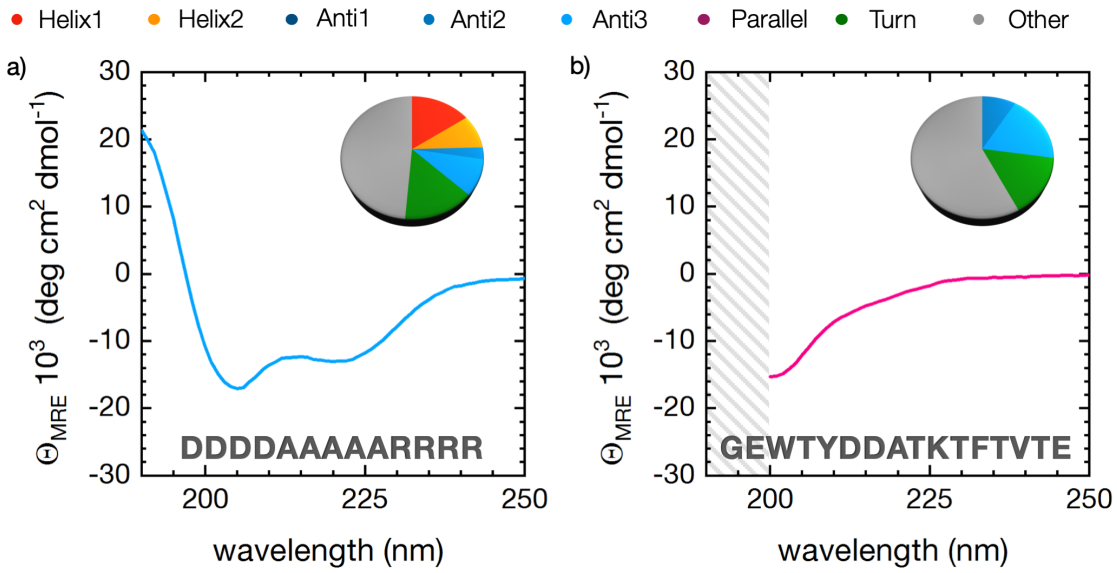


Figure 1: Experimental CD spectra of 4DAR5, dissolved in ddH_2O (a), and GB1 dissolved in a 3 mM PBS solution (b). The amount of secondary structure elements obtained by spectral analysis is reported as pie charts in the insets, together with the primary sequences. The shaded area in the GB1 spectrum indicates a region of limited accessibility due to the prominent absorbance of the buffer solution.

to the capped versions of GB1. The spectrum shows a minimum at 200 nm and a shoulder ranging from 210 nm to 230 nm. The wavelength region below 200 nm is not accessible, because the poor solubility of the peptide in pure water makes the usage of phosphate buffer saline (PBS) necessary, resulting in too strong absorption in this spectral region.[51] BeStSel predicts the macrostate secondary structure to comprise of two different anti-parallel β-sheet components (27.2 ± 1.4 % in total), a 'turn' component (16.2 ± 1.4 %) and unordered 'other' components (c.f. Figure 1 b)). Indeed, NMR experiments on GB1 in solution also reveal that the peptide is only 30 % folded in water at 25° C.[22]

De novo structural predictions

We now evaluate the intrinsic conformational propensity of 4DAR5 and GB1 with the PepFold *de novo* secondary structure prediction server.[34–36]

All five most-probable structures for 4DAR5 (labelled PF1 to PF5) present a helical con-

formation with different extent of disorder in their termini (Figure 2 (a)). For all structures, we compute the corresponding CD spectra using DichroCalc and three different parameter sets, as described in the Methods section (Figure 2 (b)). The semi-empirical parameter set gives a very good qualitative agreement to the experimental spectrum, predicting the two minima at 206 and 222 nm for all structures. The structures in best agreement (PF1 and PF3) present a helicity value H of 65% and 60%, respectively, as computed with equation 2. The *ab initio* parameter set is not able to reproduce the experimental CD spectrum for any of the structures, predicting a single minimum at about 210 nm. The inclusion of vibrational transitions (*ab initio* + vib) improves the predictions and reveals the two minima at 208 and 222 nm. However, the (negative) signal intensities are overestimated with respect to experiment.

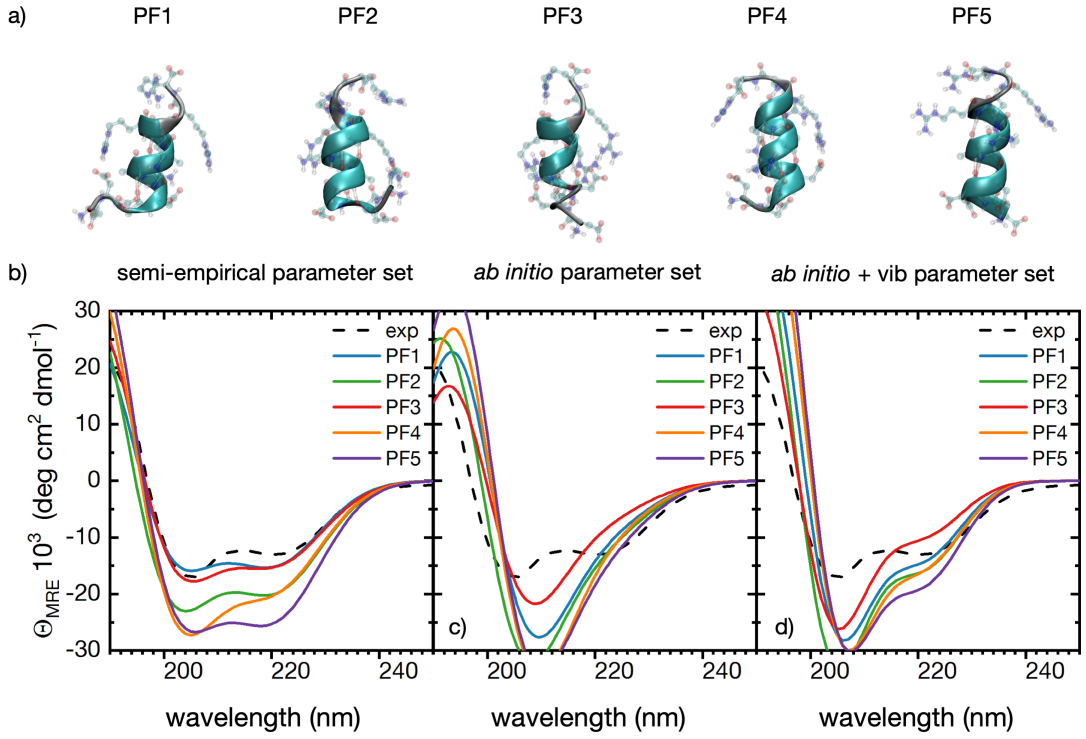


Figure 2: *De novo* predicted secondary structures and calculated CD spectra for the 4DAR5 peptide. Structure naming refers to the internal Pep-Fold classification. (a) Visualization of the predicted structures from Pep-Fold.[34–36] (b) Calculation of the associated CD spectra using three different parameter sets implemented in DichroCalc.[19]

The predicted structures for GB1 all present β-hairpin configurations, differing mostly in the number of residues within the antiparallel β-strands and in their mutual twisting (Figure 3 (a)). Irrespectively of the employed parameter set, a large variability can be observed in the corresponding DichroCalc CD spectra (Figure 3 (b)). The spectra of PF1, PF3 and PF4 present maxima in the region between 200 and 220 nm, which are entirely absent in the experiments. Only the '*ab initio* + vib' parameter set predicts spectra for PF2 and PF5 that follow at least qualitatively the experimental spectral shape. A further validation of the parameter sets was also performed on selected results of the enhanced sampling simulations, *vide infra* and Supporting Information.

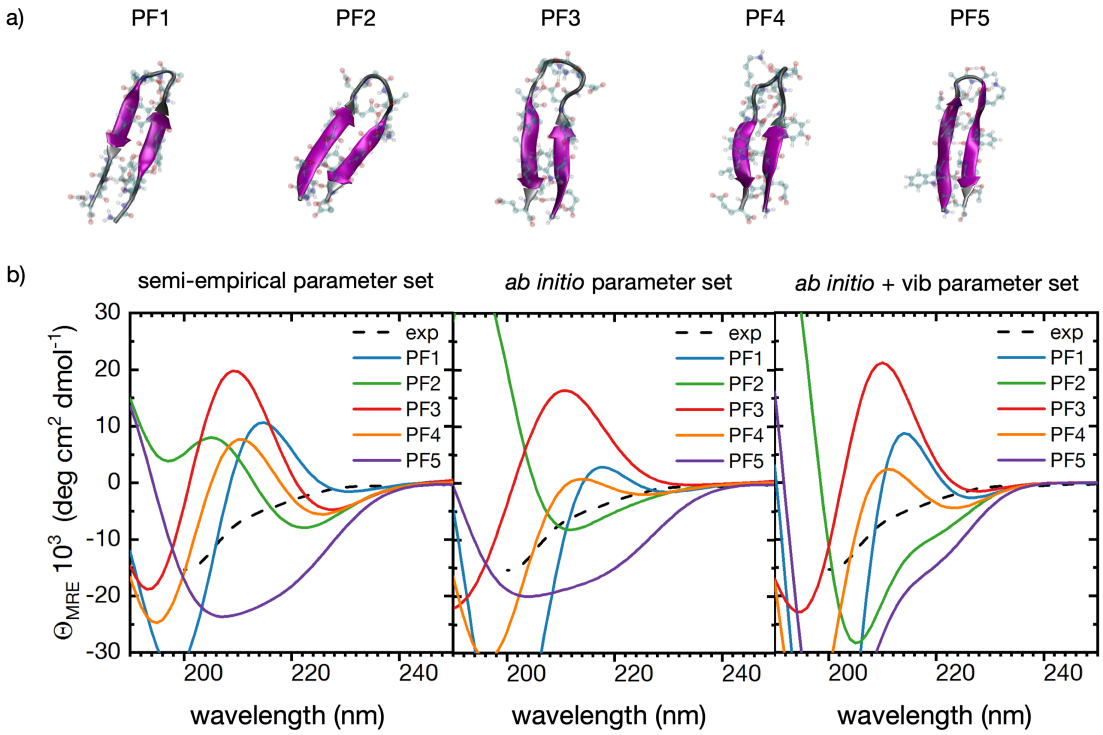


Figure 3: *De novo* predicted secondary structures and calculated CD spectra for the GB1 peptide. Structure naming refers to the internal Pep-Fold classification. (a) Visualization of the predicted structures from Pep-Fold[34–36]. (b) Calculation of the associated CD spectra using three different parameter sets implemented in DichroCalc.[19]

Enhanced sampling simulations

The whole conformational ensembles of the two peptides are explored with RESTmetaD, following our previous work.[14] The free-energy profile associated with the folding/unfolding of the 4DAR5 peptide in TIP3P water along the collective variable H is reported in Figure 4 (top panel). As already discussed in Ref.,[14] and also obtained before for the folding/unfolding of the C-terminal helix of chymotrypsin,[15] the shallow profile presents a series of 12 dicrete local minima with depths of about 0.8 kcal mol^{-1}, corresponding to different numbers of amino acids in the folded conformation, as an effect of the helicity CV definition (Eq. 2). The global profile's minimum is located at about $H = 65\,\%$. Boltzmann integration of the entire profile predicts an average value of helicity $H = 63.5\,\%$, which agrees very well with the prediction of the *de novo* PepFold server.

Following the established literature, the RESTmetaD simulations of GB1 in pure water are performed biasing the dynamics along both the radius of gyration R_{Gyr} (Eq. 3) and the number of interstrand hydrogen bonds N_H (Eq. 4).

The resulting two-dimensional free energy profile is reported in Figure 5, top right. It shows a shallow triangularly shaped basin, characterized by three main minima. The first minimum is located in an elongated valley region around $N_H \simeq 0$, spanning the R_{Gyr} range between 5 and 12 Å. This region, together with the second minimum located nearby (around $N_H \simeq 1$ and $R_{Gyr} \simeq 6$ Å), corresponds to a molten state dominated by partially folded α-helical structures (*vide infra*). The third minimum, corresponding to the native β-hairpin conformation, is located in a broader region with $3 < N_H < 5$ and $R_{Gyr} \simeq 7$. It is important to recall at this point that the relative free energy values of these three minima are known to depend strongly on the choice of the force field employed in the simulations.[23,24] The Amber03 force field, employed here for consistency with the 4DAR5 simulations, overestimates α-helices over other secondary structures,[52] which causes the free energy landscape of GB1 to be deeper at $N_H < 1$ than in the β-hairpin region, where the global minimum should in fact be located.[22] Nevertheless, this known and expected inconsistency does not hinder our

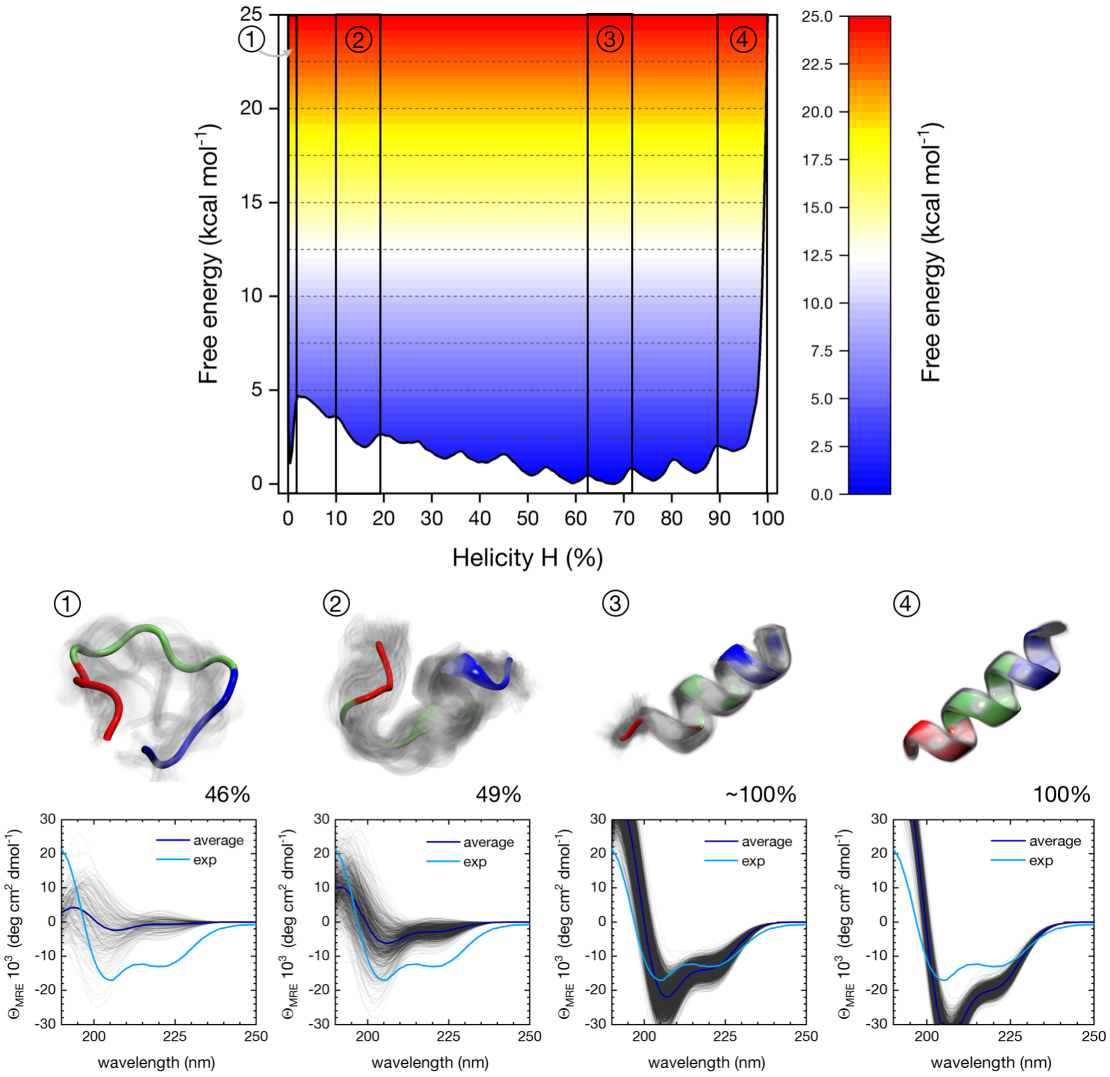


Figure 4: The free energy landscape of 4DAR5. Top: RESTmetaD free energy profile of 4DAR5 in solution with respect to the peptide's helicity H (Eq. 2). The black vertical lines delimit the regions that were considered for the subsequent cluster analysis. Bottom: visualization of the most populated clusters within the selected regions, accompanied by the corresponding computed CD spectra. The central structure of each cluster is drawn as a solid ribbon, colored according to Figure 1, while all other structures are superimposed in a transparent view. Percentages indicate the number of frames in the selected region that belong to the cluster. The CD spectra report both the computed spectra from the cluster (black) and their average (dark blue). The experimental spectrum (cyan) is also included for comparison.

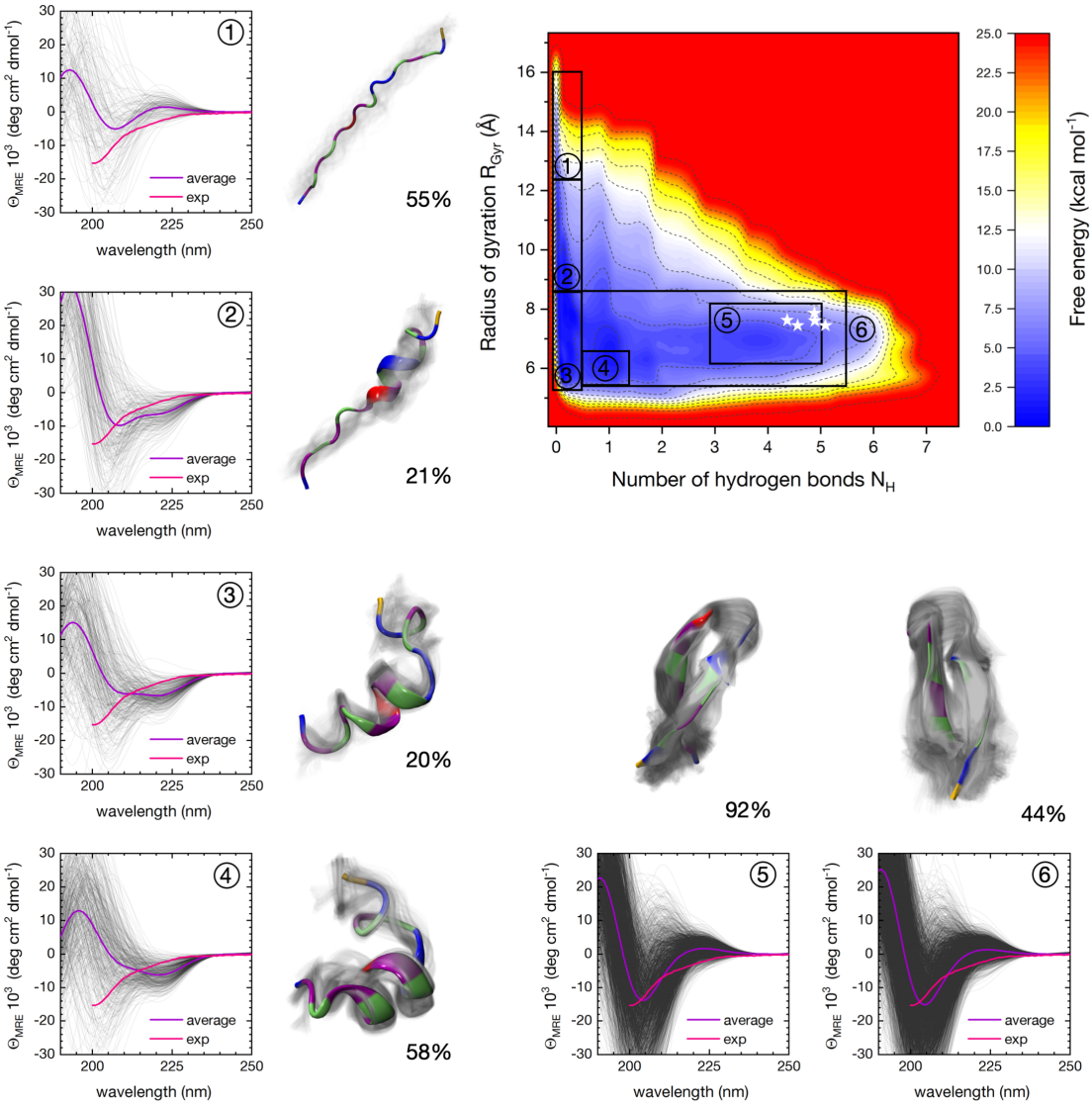


Figure 5: The free energy landscape of GB1. Top-right: two-dimensional RESTmetaD free energy surface of GB1 in solution with respect to both the radius of gyration R_{Gyr} (Eq. 3) and the number of hydrogen bonds N_H (Eq. 4). The black rectangles delimit the regions that were considered for the subsequent cluster analysis. White stars represent the values of R_{Gyr} and N_H calculated for the five *de novo* predicted structures of Figure 3 a. Surrounding the profile: visualization of the most populated clusters within the selected regions, accompanied by the corresponding computed CD spectra. The central structure of each cluster is drawn as a solid ribbon, colored according to Figure 1, while all other structures are superimposed in a transparent view. Percentages indicate the number of frames in the selected region that belong to the cluster. The CD spectra report both the computed spectra from the cluster (black) and their average (purple). The experimental spectrum (magenta) is also included for comparison.

further analysis, in which CD spectra pertinent to ensembles of conformers constrained in local regions of the sampled phase space are calculated.

Structural cluster analysis and calculation of CD spectra

The RESTmetaD ensembles presented above are now used to compute averaged CD spectra, which are compared with the experimental ones. In the present work, rather than considering the entire sampled phase space, we first select constrained regions of interest (delimited by black lines in Figures 4 and 5). We then apply a clustering algorithm based on the RMSD of backbone positions (see Methods) to all sampled microstates within each region, identify the most populated clusters, and then compute the averaged CD spectra pertinent to such clusters. To check for issues with the chosen clustering approach, we performed a clustering on selected regions of 4DAR5 and GB1 based on the RMSD between inter-residue contact maps. An overlap between the central structures of the most populated clusters obtained with the two norms reveals a very good agreement (Figure S3 of the Supporting Information). The calculations are performed only with the '*ab initio* + vib' parameter set, whose predicted spectra associated with the *de novo* structures were in good agreement with experiments both for 4DAR5 and GB1 (see above). A further verification, employing the different parameter sets available in DichroCalc, was also performed on selected regions (vide infra). Additionally, a STRIDE secondary structure determination is performed on the same clusters, to help assign the dominant conformation in each region of the CV space depending on the sequence position of each individual amino acid residues in the oligopeptides. The results are reported in Figure 6.

For 4DAR5, four regions are chosen at increasing helicity values (Figure 4). In each of the regions, the superposition of all backbone conformers in the most populated clusters are shown with the central structure highlighted in color, together with the percentage of conformers belonging to the cluster. As expected, with increasing helicity the conformational order of the ensemble grows accordingly. A completely unfolded helix corresponds to 0 %

$< H < 2\,\%$ (region 1) and a partially folded helix to $10\,\% < H < 19\,\%$ (region 2). Up to this point, the most populated clusters contain less than 50 % of the total frames in each region. Around the global free energy minimum at $63\,\% < H < 72\,\%$ (region 3), the cluster comprises almost all conformers and reveal a very high degree of helical order. The order is almost completed at $89\,\% < H < 100\,\%$ (region 4), *i.e.* leading to a fully folded helix. From the visualization of the central structures, it can be inferred how the folding of the helix seems to proceed from the N-terminus, starting from a horseshoe-like unfolded conformation. A similar conclusion can be obtained by looking at the evolution of the dihedral angle distribution in the Ramachandran plots at increasing helicity, where the C-terminal residues are the last to enter the α-helical region (Figure 6 (a)).

The computed CD spectra for each cluster reveal a correlation between degree of helicity and spectral variability among the clustered structures. A wide range of spectra is obtained within the most populated cluster of region 1, displaying, for instance, both minima and maxima around 200 nm. Going from region 2 to 4, the spectra within the most populated clusters converge towards spectral signatures associated with α-helices. As regard the comparison with the experiment, it is notable how the best agreement is obtained with the most populated cluster of region 3 (the free energy minimum), while the spectra of region 4 largely overshoot the experimental signal intensities. The deconvolution of the averaged spectrum of the most populated cluster of region 3 *via* BeStSel gives a α-helical composition of $39.6 \pm 1.4\,\%$ and a $42.2 \pm 1.4\,\%$ contribution of 'other', minor contributions are found of 'turn' and anti-parallel β-sheets. For this region we also compared the influence of the different parameter sets implemented in DichroCalc, which can be found in Figure S2 of the Supporting Information. In comparison with the STRIDE analysis and the associated Ramachandran plots (c.f. Figure 6 a)), we find around 30 % of the amino acids reveal a partially unordered conformation according to the STRIDE analysis and the presence of amino acids with backbone dihedrals in a β-sheet conformation. However, it is important to remember that the conformational/structural description of CD spectra based solely on

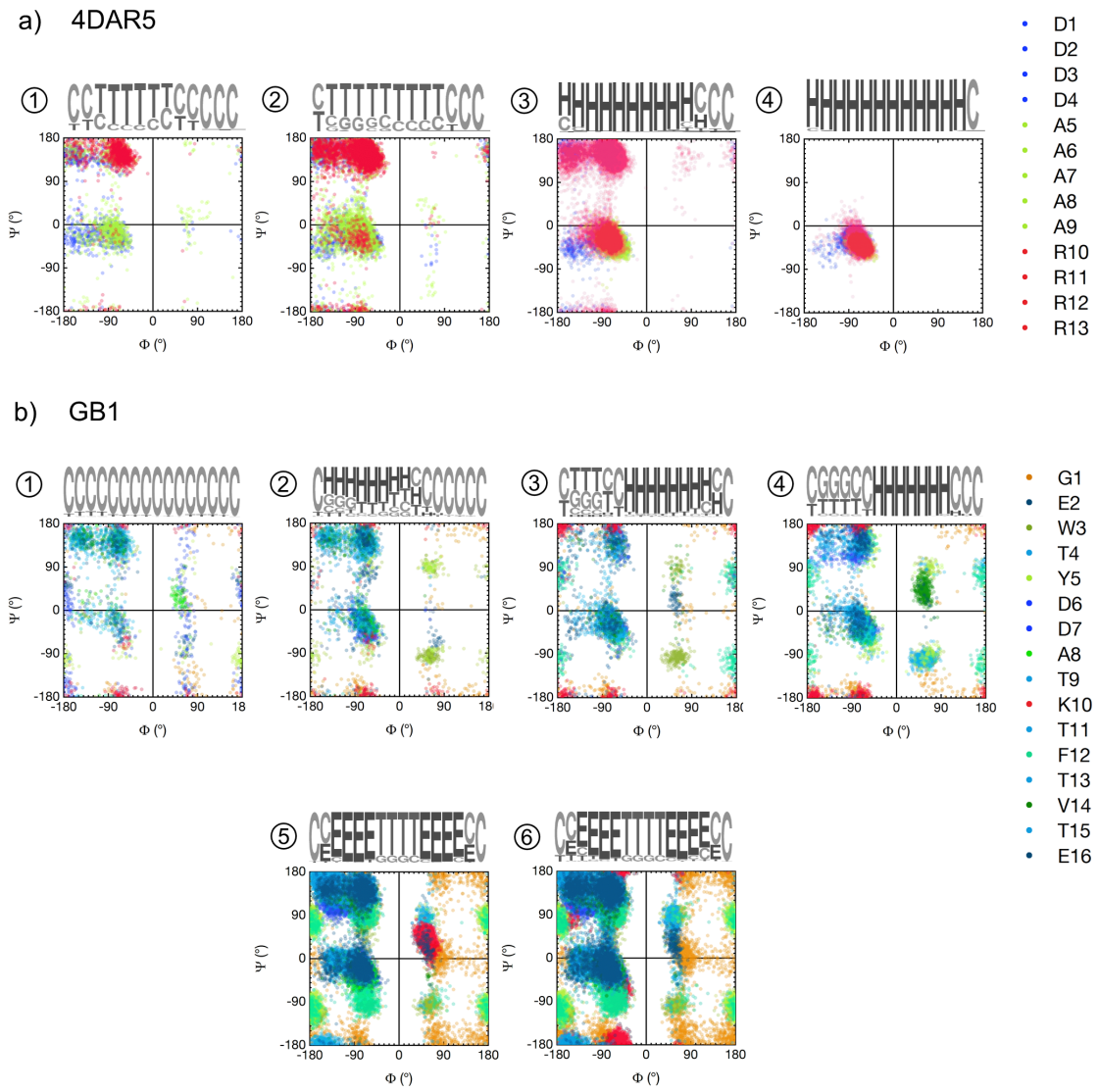


Figure 6: Ramachandran plots for the most populated clusters of a) 4DAR5 and b) GB1, in each selected region of their free energy landscapes, accompanied by a WebLogo representation[53] of the secondary structure components (from a STRIDE analysis): C, coil; T, turn; H, α-helix; G, 3-10 helix; E, extended strand in parallel or anti-parallel β-sheets.

backbone dihedrals is not sufficient.

Five regions were selected in the phase space of GB1, along the path of minimal free energy (Figure 5): 1) $0 < N_H < 0.5$ and $12.3 < R_{Gyr} < 16.0$; 2) $0 < N_H < 0.5$ and $8.6 < R_{Gyr} < 12.3$; 3) $0 < N_H < 0.5$ and $5.3 < R_{Gyr} < 8.6$; 4) $0.5 < N_H < 1.5$ and $5.4 < R_{Gyr} < 6.5$; 5) $3.0 < N_H < 5.0$ and $6.0 < R_{Gyr} < 8.0$. An additional sixth region ($0.5 < N_H < 5.4$ and $5.4 < R_{Gyr} < 8.6$) encompasses the whole free energy valley at low R_{Gyr}. Regions 1 to 3 follow the decrease in R_{Gyr} from an extended structure to a partially folded state within the minimum free energy valley at low N_H. Regions 4 and 5 include the free energy minima at increasing N_H, with region 5 corresponding to the folded β-hairpin conformation.

Strong structural clustering ($> 50\,\%$ frames in the most populated cluster) is obtained for regions 1, 4 and 5, while larger conformational variability characterize the remaining selected areas. As already anticipated above, the central cluster structures in regions 2, 3 and 4 have a significant portion of residues in an α-helical conformation, at the N-terminus in region 2 and the C-terminus in regions 3 and 4. In region 5, almost all conformers ($92\,\%$) are tightly grouped together around the expected β-hairpin structure. A β-hairpin conformation is also the central cluster structure in region 6, although in this case only $44\,\%$ of the frames belong to the most populated cluster. The angle distributions in the corresponding Ramachandran plots (c.f. Figure 6 b)) confirm the STRIDE secondary structure assignments. However, at variance with 4DAR5, the residues are spread in a wider Φ/Ψ range. Noteworthy, only in the most populated cluster of region 5 the K10 residue of GB1 adopts torsional angles characteristic of the α_L region of the Ramachandran plot, which is usually populated by residues in a turn conformation.[23] Inter-residue average contact maps for the most populated clusters in each region of the GB1 free energy landscape (Figure S4 in Supporting Information) also confirm the previous structural assignment, with typical α-helical signatures in regions 2, 3 and 4 and β-hairpin signatures in regions 5 and 6.

The observed conformational richness within the computed free energy landscape makes the GB1 system an interesting playground to test the predictive capability of the DichroCalc

parameter sets. The average spectrum corresponding to the most extended and energetically unfavored conformation (region 1) is qualitatively not dissimilar from the average spectrum of unfolded 4DAR5, although with larger spectral intensities. The spectra in regions 2, 3 and 4 do reveal α-helical signatures (e.g. local minima at about 220 and 225 nm), in agreement with the cluster and STRIDE analyses (Figure 5 and 6). However, the spectral variability among the single microstates within these clusters is very high, consistent with the overall poor degree of order therein. A high variability of CD spectra is found also in region 5, even though the structure of the conformers inside the most populated cluster are tightly overlaid. We note that a similar variability was observed in the computation of CD spectra for the *de novo* predicted GB1 structures (c.f. Figure 3), that, indeed, locate themselves within the same region of the CV space (stars in Figure 5). Despite this intrinsic variability, the averaged CD spectra in regions 5 and 6 reproduce relatively well the experimental trend, although a shallow maximum instead of an extended shoulder is predicted between 222 and 227 nm. Regarding the BestSel analysis, deconvolution of the averaged spectrum of the most populated cluster of region 5 gives a β-sheet anti-parallel composition of $30.5 \pm 4.5\,\%$ and a $18.5 \pm 4.5\,\%$ contribution of 'turn', in good agreement with the experiment ($27.2 \pm 1.4\,\%$ and $16.2 \pm 1.4\,\%$, respectively, as reported in Figure 1 (b)). We computed the influence of the different parameter sets available in DichroCalc for region 5 of GB1 (in Figure S2 of the Supporting Information.). The calculation of the CD spectra for the representative regions of 4DAR5 and GB1 using the different parameter sets confirms the applicability of the *ab initio* + vib parameter towards a generalized approach for CD spectra calculation beyond α-helical peptides

Discussion

Our analysis has focused on two model oligopeptides with suggested conformational propensities towards an α-helix (4DAR5) and a β-hairpin structure (GB1), respectively. In the case of

4DAR5, advanced-sampling RestMetaD simulations with the Amber03 force field combined with DichroCalc calculations are able to predict both the structural and the CD spectral features. In particular, the theoretical CD spectrum of the most-populated structural cluster around the global free energy minimum is in very good qualitative and in reasonable quantitative agreement with experiments, employing the general '*ab initio* + vib' parameter set of DichroCalc (see Figure 4). The predicted amount of helicity (about 60 %) agrees both with experimental CD studies in the literature[21] and with the *de novo* predictions of the PepFold server (see Figure 2).

Notwithstanding this agreement, the BestSel secondary structure analysis applied to either the experimental or the theoretically predicted CD spectra only assign 20 to 40 % of secondary structure elements to α-helical structures (see Figure 1 (a)). Given the simplicity of the conformational ensemble of 4DAR5, this is, at a first glance, surprising. Indeed, the structural variability of its backbone in the most populated cluster around the free energy minimum is rather limited, even though the residues close to the termini of the peptides are not considered *via* the collective variable. However, we have already noticed in our previous analysis of the same oligopeptide[14] that the CD ellipticity values for one and the same value of helicity H span a rather broad range for instance at 222 nm. This results in a wide CD spectral variability even in narrow regions of the conformational space (such as region 3 in Figure 4). From a structural point of view, we can interpret our results in the following way. The rather shallow shape of the free energy landscape along H allows for pronounced thermal fluctuation of the oligopeptide backbone, while still keeping the conformers within an evidently helical structure.[54] This results in CD spectral contributions similar to those arising from the disordered structures used in the reference basis set corresponding to 'other' secondary structure elements in BeStSel. In other words, we suggest that the CD response of solvated oligopeptides is very sensitive to the local conformational ordering, even among conformers presenting small RMSD variations from each other, thus resulting in a large spectral variability despite a limited structural variability.

This situation is even more evident in the case of the GB1 peptide. For instance, although the five most probable structures predicted by the *de novo* server all present a characteristic β-hairpin shape (see Figure 3 (a)), their computed CD spectra vary by a very large extent (see Figure 3 (b)) irrespectively of the employed DichroCalc parameter set. This finding confirms the sensitivity of CD spectra towards minor differences in β-sheet conformations. A similar picture emerges from analysis of the CD spectra associated will all ensemble microstates within the β-hairpin region of the free-energy landscape of GB1 (region 5 and 6 in Figure 5). Besides a shift in the minimum position, the larger disagreement is found in the wavelength region between 223 and 227 nm (c.f. Figure 5, CD spectra of region 5 and 6), where the experimental spectrum presents a shallow negative shoulder, whereas the theoretical spectrum presents a shallow positive maximum. This possibly indicates that the conformational ensemble of GB1 comprises not entirely of microstates with β-hairpin propensity, but also of a substantial amount of microstates with (partially) unordered conformers. This is indeed in agreement with experimental NMR investigations, proposing that GB1 is only $30,\%$ folded as a β-hairpin in water.[22] Indeed the STRIDE analysis performed in this regions assigns 4 out of 16 residues partially to a random coil conformation and 4 out of 16 to a turn conformation. In terms of CD reference spectra, both of these conformations would be charaterised by positive values in the wavelength region between 222 and 227 nm.

As a matter of fact, the BeStSel basis set is constructed from CD spectra of larger proteins, whose characteristic secondary structure elements are well characterised by X-ray diffraction or NMR measurements.[4,5] The protein environment exerts a strong stabilizing action over the atomic positions of such structure elements, even if they occupy terminal positions,[15] which is the case of the GB1 fragment. Solvated oligopeptides, even if carrying the same sequence and possessing the same intrinsic conformational propensity in the folded state, are subjected to larger structural fluctuations, and thus, to a considerably magnified extent, to large CD spectral variability. The CD community has recognised this fact, especially for the case of β-sheets elements and their sensitivity to mutual strand twisting.[4,5,55] BeStSel,

for instance, considers an extended parameter set associated with various kinds of β-sheets, which makes its secondary structure analysis generally more robust than other approaches, especially if applied to oligopeptides.[56]

We suggest that further improvements could be achieved if the spectral data set was extended in a way that is able to capture the large thermal fluctuations of oligopeptides within confined regions of the peptide's conformational phase space. To reach this aim, the inclusion of theoretically predicted CD spectral variability in the spirit of the approach presented in this paper could be highly beneficial. Taking into account not only the intrinsic structural variability of oligopeptides but also the CD spectral variability for each representative structural motif (as identified e.g. *via* cluster analysis techniques) will be essential especially to rationalise the conformational phase space of intrinsically disordered peptides and proteins.[57–59]

A further crucial point regarding the prediction of structural propensities of oligopeptides *via* combined experimental CD analysis and theoretical exploration of their conformational phase space is the choice of the force field.[40] When no knowledge about the structure is available *a priori*, it is not guaranteed that the employed force field predicts correctly the relative energy levels of the various local minima of the free energy surface, no matter which collective variable are chosen for the analysis. For instance, the Amber03 force field employed here notoriously overestimates the stability of helical motifs, as observed in the case of GB1. As a direct comparison, however, we found that also the more recent Amber14SB force field fails to capture the correct free energy landscape for the peptide (Figure S5 in Supporting Information). Recent developments for force fields aim at an accurate description of the ensembles of intrinsically disordered proteins,[60–62] which could also be very promising for oligopeptides. However, it is very reassuring to learn that the cluster-averaged CD spectrum predicted by the '*ab initio* + vib' parameter set of DichroCalc for *confined* regions of the free energy surfaces with clear structural ordering (e.g. around the minimum corresponding to a β-hairpin fold) does agree well with the experimental measurement.

Conclusions

We have shown that conformational ensembles can be described with atomistic precision by means of Circular Dichroism spectroscopy coupled with all-atom molecular dynamics simulations based on enhanced sampling methods. In particular, the recently introduced '*ab initio* + vib' parameter set for the calculation of CD spectra within DichroCalc was proven capable of reproducing the spectral features of peptides with both helical and β-sheet intrinsic conformational propensities. This is despite the observed sensitivity of CD spectra prediction with respect to the geometry of the individual conformers composing the ensemble. The effect of this sensitivity is that *de novo* structure prediction can give reasonable results in term of computed CD spectra in the case of extremely simple conformational phase spaces, such as pure helices, while an extended ensemble of conformers already needs to be considered for β-sheets. Aiming at a generalization of the employed approach, we envisage that for preferentially unordered oligopeptides and proteins a sequence-based structure prediction with a limited amount of conformers will be even more inappropriate. This makes essential both the usage of advanced sampling simulations combined with accurate cluster analysis and the experimental acquisition of CD spectra also in the wavelength region below 190 nm, as accessible by Synchrotron techniques.

Acknowledgement

We acknowledge fruitful discussions with Carole Perry (Nottingham Trent University, UK). This work was supported by the Deutsche Forschungsgemeinschaft under grants CO 1043/17-1, GRK 2247, CO 1043/11-1. Computational resources were provided by the North-German Supercomputing-Alliance (HLRN). JK and AM were supported by the National Research, Development and Innovation Office, Hungary under grants K120391, KH125597, and 2017-1.2.1-NKP-2017-00002. AM is a Bolyai János Research Fellow of the Hungarian Academy of Sciences.

References

(1) Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R. M.; Wu, K.-P.; Baum, J. Distinguishing among Structural Ensembles of the GB1 Peptide: REMD Simulations and NMR experiments. *J. Am. Chem. Soc.* **2007**, *129*, 4858–4859, DOI: 10.1021/ja0677517, PMID: 17402734.

(2) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884, DOI: 10.1038/nature02261.

(3) Kardos, J.; Micsonai, A.; Pál-Gábor, H.; Petrik, É.; Gráf, L.; Kovács, J.; Lee, Y.-H.; Naiki, H.; Goto, Y. Reversible Heat-Induced Dissociation of β2-Microglobulin Amyloid Fibrils. *Biochemistry* **2011**, *50*, 3211–3220, DOI: 10.1021/bi2000017.

(4) Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E3095–E3103, DOI: 10.1073/pnas.1500851112.

(5) Micsonai, A.; Bulyáki, É.; Moussong, É.; Kun, J.; Kardos, J.; Wien, F.; Réfrégiers, M.; Lee, Y.-H.; Goto, Y. BeStSel: a Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra. *Nucleic Acids Res.* **2018**, *46*, W315–W322, DOI: 10.1093/nar/gky497.

(6) Whitmore, L.; Wallace, B. A. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* **2008**, *89*, 392–400, DOI: 10.1002/bip.20853.

(7) Belton, D. J.; Plowright, R.; Kaplan, D. L.; Perry, C. C. A Robust Spectroscopic Method for the Determination of Protein Conformational Composition - Application to the Annealing of Silk. *Acta Biomater.* **2018**, *73*, 355 – 364, DOI: 10.1016/j.actbio.2018.03.058.

(8) Lopes, J. L. S.; Miles, A. J.; Whitmore, L.; Wallace, B. A. Distinct Circular Dichroism Spectroscopic Signatures of Polyproline II and Unordered Secondary Structures: Applications in Secondary Structure Analyses. *Protein Sci.* **2014**, *23*, 1765–1772, DOI: 10.1002/pro.2558.

(9) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872 – 877, DOI: 10.1016/j.bbagen.2014.10.019, Recent developments of molecular dynamics.

(10) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: a More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438, DOI: 10.1021/jp204407d.

(11) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603, DOI: 10.1103/physrevlett.100.020603.

(12) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441, DOI: 10.1021/ja062463w.

(13) Schneider, J.; Colombi Ciacchi, L. Specific Material Recognition by Small Peptides Mediated by the Interfacial Solvent Structure. *J. Am. Chem. Soc.* **2012**, *134*, 2407–2413, DOI: 10.1021/ja210744g.

(14) Meißner, R. H.; Schneider, J.; Schiffels, P.; Colombi Ciacchi, L. Computational Prediction of Circular Dichroism Spectra and Quantification of Helicity Loss upon Peptide Adsorption on Silica. *Langmuir* **2014**, *30*, 3487–3494, DOI: 10.1021/la500285m.

(15) Hildebrand, N.; Michaelis, M.; Wurzler, N.; Li, Z.; Hirst, J. D.; Micsonai, A.; Kardos, J.; Gil-Ley, A.; Bussi, G.; Köppen, S. et al. Atomistic Details of Chymotrypsin Conforma-

tional Changes upon Adsorption on Silica. *ACS Biomaterials Science & Engineering* **2018**, *4*, 4036–4050, DOI: 10.1021/acsbiomaterials.8b00819.

(16) Goerigk, L.; Grimme, S. Calculation of Electronic Circular Dichroism Spectra with Time-Dependent Double-Hybrid Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 767–776, DOI: 10.1021/jp807366r, PMID: 19102628.

(17) Bulheller, B. M.; Hirst, J. D. DichroCalc - Circular and Linear Dichroism Online. *Bioinformatics* **2009**, *25*, 539–540, DOI: 10.1093/bioinformatics/btp016.

(18) Ianeselli, A.; Orioli, S.; Spagnolli, G.; Faccioli, P.; Cupellini, L.; Jurinovich, S.; Mennucci, B. Atomic Detail of Protein Folding Revealed by an Ab Initio Reappraisal of Circular Dichroism. *J. Am. Chem. Soc.* **2018**, *140*, 3674–3682, DOI: 10.1021/jacs.7b12399, PMID: 29473417.

(19) Bulheller, B. M.; Rodger, A.; Hirst, J. D. Circular and Linear Dichroism of Proteins. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2020–2035, DOI: 10.1039/B615870F.

(20) Read, M. J.; Burkett, S. L. Asymmetric α-Helicity Loss within a Peptide Adsorbed onto Charged Colloidal Substrates. *J. Colloid Interface Sci.* **2003**, *261*, 255 – 263, DOI: 10.1016/S0021-9797(03)00092-4.

(21) Burkett, S. L.; Read, M. J. Adsorption-Induced Conformational Changes of α-Helical Peptides. *Langmuir* **2001**, *17*, 5059–5065.

(22) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. Enhanced Hairpin Stabilty through Loop Design: The Case of the Protein GB1 Domain Hairpin. *J. Am. Chem. Soc.* **2004**, *126*, 7238–7243, DOI: 10.1021/ja0379520, PMID: 15186161.

(23) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins* **2011**, *79*, 1318–1328, DOI: 10.1002/prot.22972.

(24) Hazel, A. J.; Walters, E. T.; Rowley, C. N.; Gumbart, J. C. Folding Free Energy Landscapes of β-Sheets with Non-Polarizable and Polarizable CHARMM Force Fields. *J. Chem. Phys.* **2018**, *149*, 072317, DOI: 10.1063/1.5025951.

(25) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. The Unfolded Ensemble and Folding Mechanism of the C-Terminal GB1 β-Hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 13938–13944, DOI: 10.1021/ja803652f.

(26) Soranno, A.; Cabassi, F.; Orselli, E.; Cellmer, T.; Gori, A.; Longhi, R.; Buscaglia, M. Dynamics of Structural Elements of GB1 β-Hairpin Revealed by Tryptophan-Cysteine Contact Formation Experiments. *J. Phys. Chem. B* **2018**, *122*, 11468–11477, DOI: 10.1021/acs.jpcb.8b07399.

(27) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197–2201, DOI: 10.1021/ct900202f.

(28) Sancho, D. D.; Mittal, J.; Best, R. B. Folding Kinetics and Unfolded State Dynamics of the GB1 Hairpin from Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 1743–1753, DOI: 10.1021/ct301033r.

(29) Ahalawat, N.; Mondal, J. Assessment and Optimization of Collective Variables for Protein Conformational Landscape: GB1 β-Hairpin as a Case Study. *J. Chem. Phys.* **2018**, *149*, 094101, DOI: 10.1063/1.5041073.

(30) Evans, D. A.; Wales, D. J. Folding of the GB1 Hairpin Peptide from Discrete Path Sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090, DOI: 10.1063/1.1759317.

(31) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 nm. *Protein Sci.* **2013**, *22*, 851–858, DOI: 10.1002/pro.2253.

(32) Kuipers, B. J.; Gruppen, H. Prediction of Molar Extinction Coefficients of Proteins and Peptides Using UV Absorption of the Constituent Amino Acids at 214 nm to Enable Quantitative Reverse Phase High-Performance Liquid Chromatography-Mass Spectrometry Analysis. *J. Agric. Food Chem.* **2007**, *55*, 5445–5451, DOI: 10.1021/jf0703371.

(33) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2007**, *1*, 2876–2890, DOI: 10.1038/nprot.2006.202.

(34) Shen, Y.; Maupetit, J.; Derreumaux, P.; Tuffery, P. Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J. Chem. Theory Comput.* **2014**, *10*, 4745–4758, DOI: 10.1021/ct500592m.

(35) Thevenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tuffery, P. PEP-FOLD: an Updated De Novo Structure Prediction Server for Both Linear and Disulfide Bonded Cyclic Peptides. *Nucleic Acids Res.* **2012**, *40*, W288–W293, DOI: 10.1093/nar/gks419.

(36) Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: Faster De Novo Structure Prediction for Linear Peptides in Solution and in Complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454, DOI: 10.1093/nar/gkw329.

(37) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19, DOI: 10.1006/jcph.1995.1039.

(38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1996**, *118*, 2309–2309, DOI: 10.1021/ja00124a002.

(39) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. et al. A Point-Charge Force Field for Molecular Mechanics

Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012, DOI: 10.1002/jcc.10349.

(40) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, DOI: 10.1021/acs.jctc.5b00255.

(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33–38, DOI: 10.1016/0263-7855(96)00018-5.

(42) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217–228, DOI: 10.1021/ct050250b.

(43) Bussi, G. Hamiltonian Replica Exchange in GROMACS: a Flexible Implementation. *Mol. Phys* **2013**, *112*, 379–384, DOI: 10.1080/00268976.2013.824126.

(44) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun* **2009**, *180*, 1961–1972, DOI: 10.1016/j.cpc.2009.05.011.

(45) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem.* **1999**, *38*, 236–240, DOI: 10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.0.co;2-m.

(46) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637, DOI: 10.1002/bip.360221211.

(47) Woody, R. W.; Sreerama, N. Comment on "Improving Protein Circular Dichroism Calculations in the Far-Ultraviolet through Reparametrizing the Amide Chromophore". *J. Chem. Phys.* **1999**, *111*, 2844–2845, DOI: 10.1063/1.479562.

(48) Besley, N. A.; Hirst, J. D. Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* **1999**, *121*, 9636–9644, DOI: 10.1021/ja9906271.

(49) Li, Z.; Robinson, D.; Hirst, J. D. Vibronic Structure in the Far-UV Electronic Circular Dichroism Spectra of Proteins. *Faraday Discuss.* **2015**, *177*, 329–344, DOI: 10.1039/c4fd00163j.

(50) Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. Stabilizing Capping Motif for β-Hairpins and Sheets. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10466–10471, DOI: 10.1073/pnas.0913534107.

(51) Kelly, S. M.; Jess, T. J.; Price, N. C. How to Study Proteins by Circular Dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* **2005**, *1751*, 119 – 139, DOI: 10.1016/j.bbapap.2005.06.005.

(52) Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* **2008**, *95*, L07–L09, DOI: 10.1529/biophysj.108.132696.

(53) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190, DOI: 10.1101/gr.849004.

(54) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. The Concept of a Random Coil: Residual Structure in Peptides and Denatured Proteins. *Fold Des.* **1996**, *1*, R95–R106, DOI: 10.1016/S1359-0278(96)00046-6.

(55) Daura, X.; Bakowies, D.; Seebach, D.; Fleischhauer, J.; van Gunsteren, W. F.; Krüger, P. Circular dichroism Spectra of β-Peptides: Sensitivity to Molecular Struc-

ture and Effects of Motional Averaging. *Eur. Biophys. J.* **2003**, *32*, 661–670, DOI: 10.1007/s00249-003-0303-1.

(56) Sola-Rabada, A.; Michaelis, M.; Oliver, D. J.; Roe, M. J.; Colombi Ciacchi, L.; Heinz, H.; Perry, C. C. Interactions at the Silica-Peptide Interface: Influence of the Extent of Functionalization on the Conformational Ensemble. *Langmuir* **2018**, *34*, 8255–8263, DOI: 10.1021/acs.langmuir.8b00874.

(57) Fisher, C. K.; Stultz, C. M. Constructing Ensembles for Intrinsically Disordered Proteins. *Curr Opin Struct Biol* **2011**, *21*, 426 – 431, DOI: 10.1016/j.sbi.2011.04.001.

(58) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524, DOI: 10.1021/acs.jctc.5b00736, PMID: 26574339.

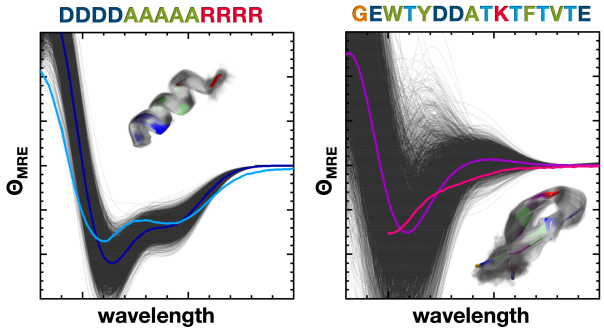
(59) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr Opin Struct Biol* **2017**, *42*, 106 – 116, DOI: 10.1016/j.sbi.2016.12.004.

(60) Song, D.; Luo, R.; Chen, H.-F. The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *J. Chem. Inf. Model.* **2017**, *57*, 1166–1178, DOI: 10.1021/acs.jcim.7b00135, PMID: 28448138.

(61) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71, DOI: 10.1038/nmeth.4067.

(62) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E4758–E4766, DOI: 10.1073/pnas.1800690115.

TOC Graphic


DDDDAAAAARRRR
GEWTYDDATKTFTVTE
Θ MRE
wavelength
Θ MRE
wavelength