

# Force Field Effects in Simulations of Flexible Peptides with Varying Polyproline II Propensity

Stéphanie Jephthah, Francesco Pesce, Kresten Lindorff-Larsen,\* and Marie Skepö\*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 6634–6646



Read Online

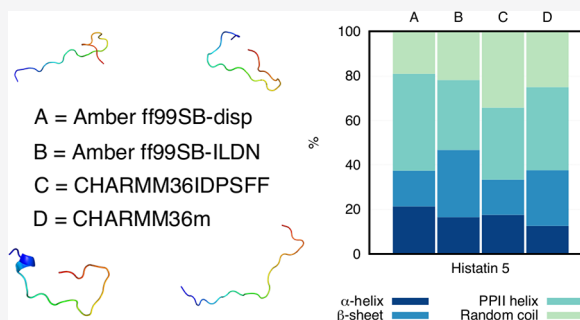
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Five peptides previously suggested to possess polyproline II (PPII) structure have here been investigated by using atomistic molecular dynamics simulations to compare how well four different force fields known for simulating intrinsically disordered proteins relatively well (Amber ff99SB-disp, Amber ff99SB-ILDN, CHARMM36IDPSFF, and CHARMM36m) can capture this secondary structure element. The results revealed that all force fields sample PPII structures but to different extents and with different propensities toward other secondary structure elements, in particular, the  $\beta$ -sheet and “random coils”. A cluster analysis of the simulations of histatin 5 also revealed that the conformational ensembles of the force fields are quite different. We compared the simulations to circular dichroism and nuclear magnetic resonance spectroscopy experiments and conclude that further experiments and methods for interpreting them are needed to assess the accuracy of force fields in determining PPII structure.



## 1. INTRODUCTION

Intrinsically disordered proteins and regions (IDPs and IDRs)—also recognized as natively unfolded proteins and peptides—are characterized by the lack of a well-defined tertiary structure in aqueous solution. Their structural properties are known to vary significantly, which makes it difficult to study them by standard methods. For example, because of their flexible behavior, IDPs and IDRs cannot be crystallized,<sup>1</sup> and until relatively recently, many molecular simulations of flexible peptides and proteins had a strong bias for sampling  $\alpha$ -helical and  $\beta$ -sheet structures as well as too compact and stable structures.<sup>2–4</sup>

Despite IDPs often being thought of as “unordered”, studies originating from the 1970s discovered that a few natively unfolded peptides possessed some degree of local order in their backbones, identified as the left-handed polyproline II (PPII) helix.<sup>5</sup> Although the PPII helix was discovered a long time ago, it is still substantially less known than, for example, the  $\alpha$ -helix and the  $\beta$ -sheet.<sup>6</sup> One of the more recognized occurrences of the PPII helix might be as being part of the triple-helix structure of collagen, where it helps stabilize the collagen structure,<sup>7</sup> and indeed specific efforts have been devoted to optimizing computational models for collagen.<sup>8,9</sup> Another well-known and important property of the PPII helix has been observed in the binding to SH3 domains,<sup>10</sup> where it facilitates and mediates protein–protein interactions.<sup>11,12</sup>

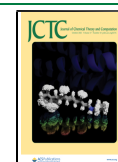
As a secondary structure element, the PPII helix is decidedly different from the  $\alpha$ -helix and the  $\beta$ -sheet and perhaps less well-known, although frequently occurring in many proteins. The PPII helix has backbone dihedral angles of approximately  $(\phi, \psi)$

$= (-75^\circ, +145^\circ)$  with a helical pitch of 9.3 Å/turn and 3.0 residues/turn, which causes it to become quite extended.<sup>6,13,14</sup> The peptide bond in a PPII helix is always found as the *trans* isomer ( $\omega = +180^\circ$ ) since the *cis* isomer ( $\omega = 0^\circ$ ) yields a different secondary structure known as the polyproline I (PPI) helix ( $(\phi, \psi) = (-75^\circ, +160^\circ)$  with a helical pitch of 5.6 Å/turn and 3.3 residues/turn).<sup>15</sup>

One popular experimental technique for determining the secondary structure of proteins is circular dichroism (CD) spectroscopy. In the CD spectrum, the PPII helix is often associated with a strong band with negative ellipticity around 198 nm and a weak positive band around 218 nm.<sup>5,6,16</sup> There are several software packages available for analyzing CD data and providing estimates of the relative secondary structure content, at least in terms of  $\alpha$ -helices and  $\beta$ -sheets.<sup>17–19</sup> Unfortunately, these algorithms may fall short when it comes to analyzing spectra of more disordered proteins that contain several less common secondary structure elements, including the PPII helix, and where the structural elements are not fixed in time. In such cases, the remaining secondary structure elements are lumped together and categorized as “others” or “random coils”. A similar problem is also encountered when the secondary structure

Received: April 23, 2021

Published: September 15, 2021



content of protein structures is determined from simulations. Although the widely used DSSP (Dictionary of Secondary Structure of Proteins) program is able to identify and quantify a wider collection of secondary structure elements, it does not include the PPII helix. Fortunately, there is other software available that utilizes modified DSSP assignment to also include the PPII structure.<sup>20,21</sup>

The development of force fields for simulating IDPs is constantly evolving to help alleviate the problem of overly collapsed structures in simulations and to make the simulations as computationally efficient and accurate as possible.<sup>2–4</sup> So far in force field development, the focus has mainly been on optimizing two aspects of IDP simulations.<sup>22–25</sup> The first aspect is the secondary structure propensities, which are often modified by adjusting the protein backbone dihedral parameters.<sup>26–32</sup> The second aspect concerns the balance of the protein–solvent interactions,<sup>33,34</sup> which is crucial to not sampling too compact IDP conformations and to accurately capturing the more extended conformations. This is generally controlled by increasing and fine-tuning the interaction between the protein and the water in the simulations.

To evaluate new force fields for simulations of flexible peptides and proteins, different properties can be considered. Nuclear magnetic resonance (NMR) observables, such as scalar couplings and chemical shifts, are used for assessing force field accuracy by comparing simulated and experimental values and are in particular sensitive to local structural properties. Comparisons of scattering curves and the radius of gyration are used to evaluate the global compactness of the simulated proteins. In addition to such direct comparison to experimental observables, secondary structure propensities are often also assessed, and although more comprehensive analyses sometimes are used, they are most often restricted to the  $\alpha$ -helix, the  $\beta$ -sheet, and the “random coil”. Conformational clustering is also sometimes used as a tool in force field analyses.

Here we present a study where five different short peptides (7–24 residues long) with varying PPII propensities, as well as five variants of one of the peptides, have been simulated with four different force fields that are known to work relatively well for simulating IDPs. Our analyses were mainly focused on differences regarding the secondary structure content across peptides and force fields. Our findings revealed that, although all the chosen force fields give rise to conformational ensembles with some level of PPII structure, they do so to different extents and with different propensities toward other secondary structure elements. Additionally, all force fields capture a trend showing that the PPII content increases with the number of Pro residues in peptide chains consisting of only Ala and Pro residues.

## 2. METHODS

**2.1. Molecular Dynamics Simulations.** Five different peptides known to possess PPII structure, as well as five different variants of one of the peptides, were simulated using atomistic molecular dynamics (MD) simulations. Names and amino acid sequences of the selected peptides are shown in Table 1.

The simulations were performed with the GROMACS package (ver. 4.6.7),<sup>35–37</sup> with four different force fields that have also previously been used for simulating IDPs: (A) the AMBER ff99SB-disp force field with its own TIP4P-D-type water model,<sup>38</sup> (B) the AMBER ff99SB-ILDN force field<sup>28</sup> with the TIP4P-D water model,<sup>34</sup> (C) the CHARMM36IDPSFF<sup>39,40</sup> force field with the CHARMM-modified TIP3P water model,<sup>41</sup> and (D) the CHARMM36m force field,<sup>31</sup> also with the

**Table 1. Names, Number of Amino Acid Residues (#aa), and Amino Acid Sequences of the Peptides Used in This Study<sup>a</sup>**

Name	#aa	Sequence
A7	7	AAAAAAA
P-113	12	AKRHHGYKRRKFH
P <sub>13</sub>	13	PPPPPPPPPPPPP
V1: P <sub>6</sub> AP <sub>6</sub>	13	PPPPPPAPPPPPP
V2: (PA) <sub>6</sub> P	13	PAPAPAPAPAPAP
V2.2: A <sub>6</sub> P <sub>7</sub>	13	AAAAAAPPPIPPP
V3: (P <sub>3</sub> A <sub>2</sub> ) <sub>2</sub> P <sub>3</sub>	13	PPPAAPPPAAPPP
V4: (A <sub>3</sub> P <sub>2</sub> ) <sub>2</sub> A <sub>3</sub>	13	AAAPPAAPPPAAA
Pep3	13	DRALGIP <sup>EL</sup> VNQG
Hst5	24	DSHAKRHHGYKRR <sup>F</sup> HEKHHSR <sup>RGY</sup>

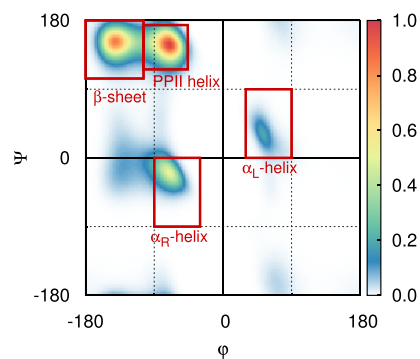
<sup>a</sup>Positively charged amino acids are shown in blue, and negatively charged amino acids are shown in red.

**Table 2. Force Field Notations**

force field	abbreviation	one-letter code
AMBER ff99SB-disp	A99SB-disp	A
AMBER ff99SB-ILDN	A99SB-ILDN	B
CHARMM36IDPSFF	C36IDPSFF	C
CHARMM36m	C36m	D

**Table 3. Force Fields (FF), Total Production Run Times (*t*), and Temperatures (*T*) of All the Simulations**

peptide(s)	FF	<i>t</i> ( $\mu$ s)	<i>T</i> (K)
A7, P-113, P <sub>13</sub> , V1, V2, V3, V4, Pep3	A	5	300
A7, P <sub>13</sub> , V1, V2, V2.2, V3, V4, Pep3	B	5	300
A7, P-113, P <sub>13</sub> , V1, V2, V3, V4, Pep3	C	5	300
A7, P-113, P <sub>13</sub> , V1, V2, V3, V4, Pep3	D	5	300
P-113	B	12	300
Hst5	A	5	293
Hst5	B	7	293
Hst5	C	5	293
Hst5	D	5	293



**Figure 1.** Example of a Ramachandran map illustrating the four different secondary structure regions analyzed in this study. The  $\beta$ -sheet region is defined by  $-180^\circ \leq \phi < -104^\circ$  and  $180^\circ \leq \psi \leq 104^\circ$ , the PPII helix region is defined by  $-104^\circ \leq \phi < -46^\circ$  and  $116^\circ \leq \psi \leq 174^\circ$ , the  $\alpha_R$ -helix region is defined by  $-90^\circ \leq \phi \leq -30^\circ$  and  $-90^\circ \leq \psi \leq 0^\circ$ , and the  $\alpha_L$ -helix region is defined by  $30^\circ \leq \phi \leq 90^\circ$  and  $0^\circ \leq \psi \leq 90^\circ$ . Anything outside of these regions is classified as “random coil”. The plot is normalized for a maximum intensity of 1. The displayed angles of the example were obtained from simulations of Hst5 with A99SB-ILDN.

**Table 4.** Radius of Gyration,  $R_g$  (nm), for the Five Peptides with the Four Different Force Fields

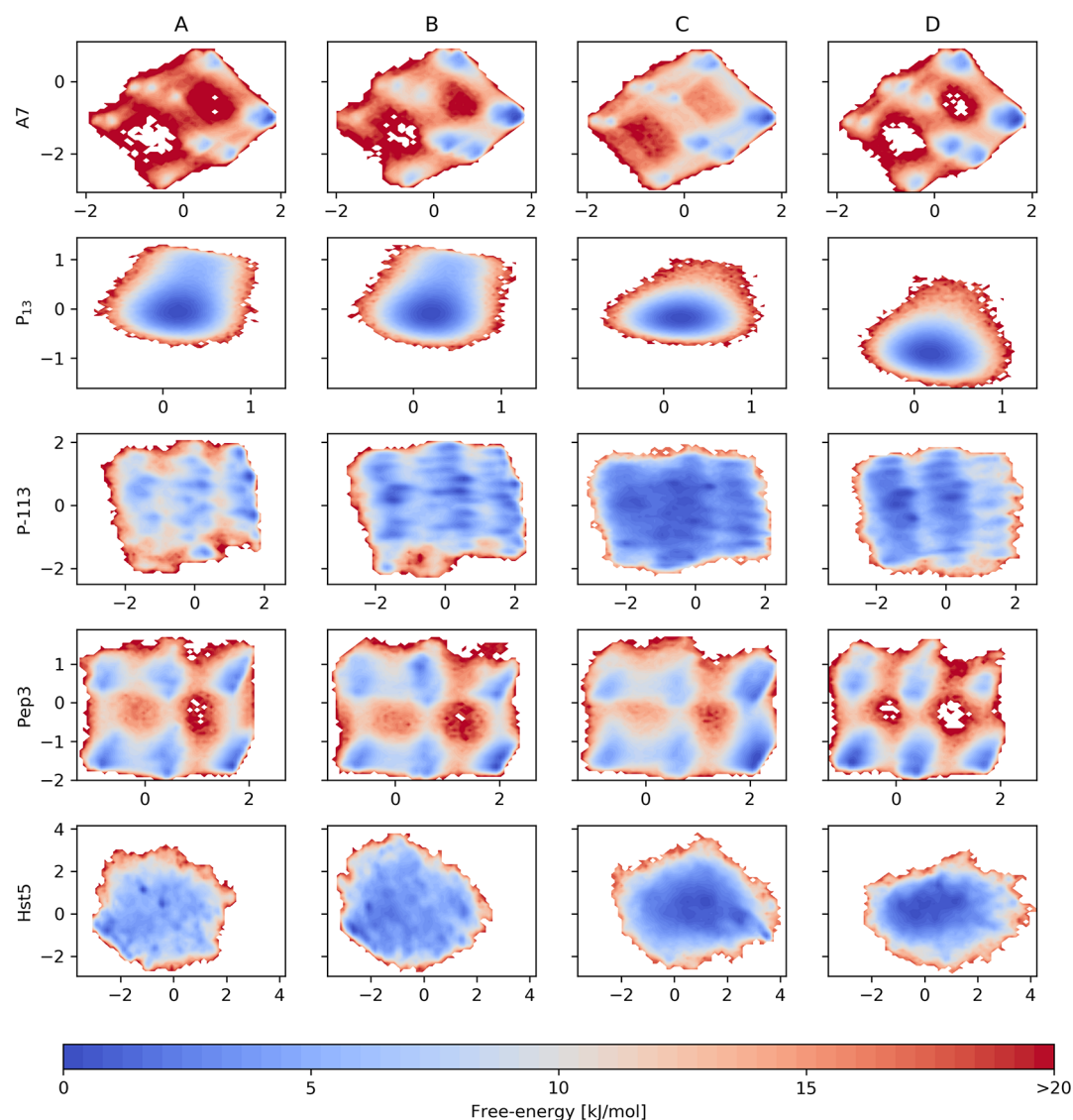
FF	A7 <sup>a</sup>	P-113	P <sub>13</sub> <sup>a</sup>	Pep3	Hst5
A	0.62	0.97 ± 0.01	1.14	0.91 ± 0.03	1.29 ± 0.08
B	0.63	0.91 ± 0.03	1.14	0.94 ± 0.01	1.29 ± 0.05
C	0.60	0.86 ± 0.01	1.13	0.87 ± 0.01	1.18 ± 0.02
D	0.61	0.92 ± 0.02	1.11	0.97 ± 0.01	1.35 ± 0.03

<sup>a</sup>The values of A7 and P<sub>13</sub> are reported without error margins because their errors are smaller than 0.005 nm.

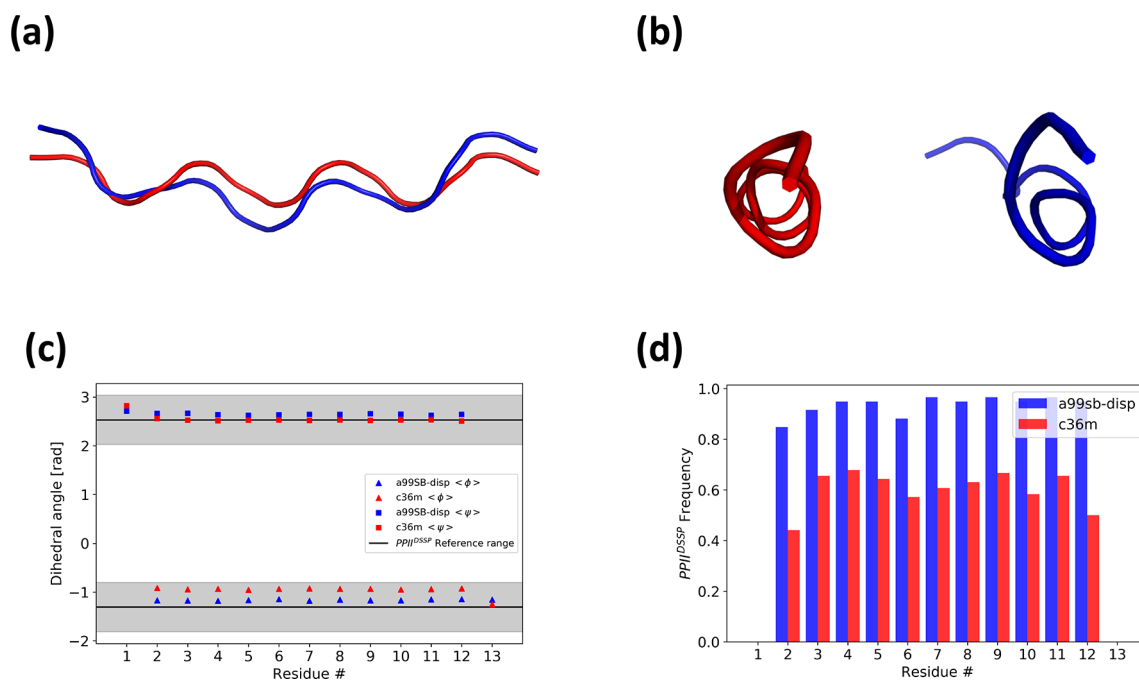
CHARMM-modified TIP3P water model. When discussed in the text, the force fields are referred to by their abbreviations, and when described in the figures and tables, they are referred to by their one-letter code (see Table 2). A rhombic dodecahedron was used as a simulation box, with periodic boundary conditions in all directions. A minimum distance of 1 nm was set between the solute and the box edges. The initial, linear protein structures were built with the use of PyMOL.<sup>42</sup>

The Verlet leapfrog algorithm,<sup>43</sup> with a time step of 2 fs, was used to integrate the equations of motion. Nonbonded interactions were computed with a Verlet list cutoff scheme, short-ranged interactions were calculated by using a pair list with a cutoff of 1 nm, and long-ranged electrostatics were evaluated by using particle mesh Ewald summation<sup>44</sup> with cubic interpolation and a grid spacing of 0.16 nm. Long-ranged dispersion interactions were applied to the energies and pressures of the simulated systems. All bond lengths were constrained by using the LINCS algorithm.<sup>45</sup> A velocity-rescaling thermostat<sup>46</sup> with a relaxation time of 0.1 ps was used to keep a temperature of 293 or 300 K (see Table 3 for details), and a Parrinello–Rahman barostat<sup>47</sup> kept the pressure at 1 bar throughout the simulations. A relaxation time of 2 ps was used, and the isothermal compressibility was set to that of water, i.e.,  $4.5 \times 10^{-5} \text{ bar}^{-1}$ .

Energy minimization was done by using the steepest descent algorithm. Equilibration of the temperature and pressure was done in two steps and with position restraints on the proteins: (1) 500 ps in the NVT ensemble and (2) 1000 ps in the NPT



**Figure 2.** Free energy surfaces as a function of the first ( $x$ -axis) and second ( $y$ -axis) principal components. As a result of performing the PCA on the aggregated trajectories, the PC coordinates for all the force fields on a row are the same.



**Figure 3.** Structural analysis of free energy minima in A99SB-disp (blue) and C36m (red). (a) Side view and (b) C-terminal view of two representative  $P_{13}$  structures from A99SB-disp and C36m. (c) Average per residues  $\phi$  and  $\psi$  dihedrals, compared to the PPII helix range defined in DSSP-PPII. (d) Per residue probability of PPII conformations as assigned by DSSP-PPII.

ensemble. Five replicates with different starting seeds were used for each simulation. The final production runs were performed in the  $NPT$  ensemble for a total of  $5 \mu s$  ( $5 \times 1 \mu s$ ) for the majority of the peptides. Hst5 with A99SB-ILDN was run for a total of  $7 \mu s$  ( $1 \times 3 \mu s + 2 \times 2 \mu s$ ), and P-113 with A99SB-ILDN was run for a total of  $12 \mu s$  ( $1 \times 8 \mu s + 2 \times 2 \mu s$ ). The differences among all the simulations are summarized in Table 3. Simulation data of Hst5 with A99SB-ILDN and Hst5 with C36m have previously been published in the paper by Jephthah et al.<sup>48</sup>

**2.2. Simulation Analyses.** The GROMACS tool `g_analyze` was used to obtain averages, autocorrelation functions, and error estimates (block averaging)<sup>49</sup> of the radius of gyration and the end-to-end distance of the simulated peptides. Results from these analyses were used to assess convergence and are presented in the Supporting Information (Figures S1–S15). The GROMACS tool `g_cluster` was used with the GROMOS algorithm<sup>50</sup> to obtain conformational clusters, and to obtain frames for representative structures. All protein structures were visualized and rendered with PyMOL.<sup>42</sup>

**2.2.1. Principal Component Analysis.** Principal component analysis (PCA) is a dimensionality reduction method that makes it possible to represent a fraction of the information contained in a large set of variables (or features) in a smaller set. This is achieved by calculating the eigenvectors (or principal components) of the variables' covariance matrix. A PCA was performed, for each peptide, on an aggregated trajectory made by concatenating the trajectories resulting from the four different force fields. This, as suggested in ref 51, ensures a robust comparison of the force fields by projecting the resulting trajectories onto common principal components. PCA calculations were performed with pyEMMA<sup>52</sup> using as features the cosine and sine of each backbone dihedral. The analyses were based on the first two principal components.

**2.2.2. Secondary Structure Analysis.** The secondary structure was analyzed by using three methods: (i) the DSSP algorithm,<sup>53</sup> (ii) the DSSP-PPII algorithm,<sup>20</sup> and (iii)

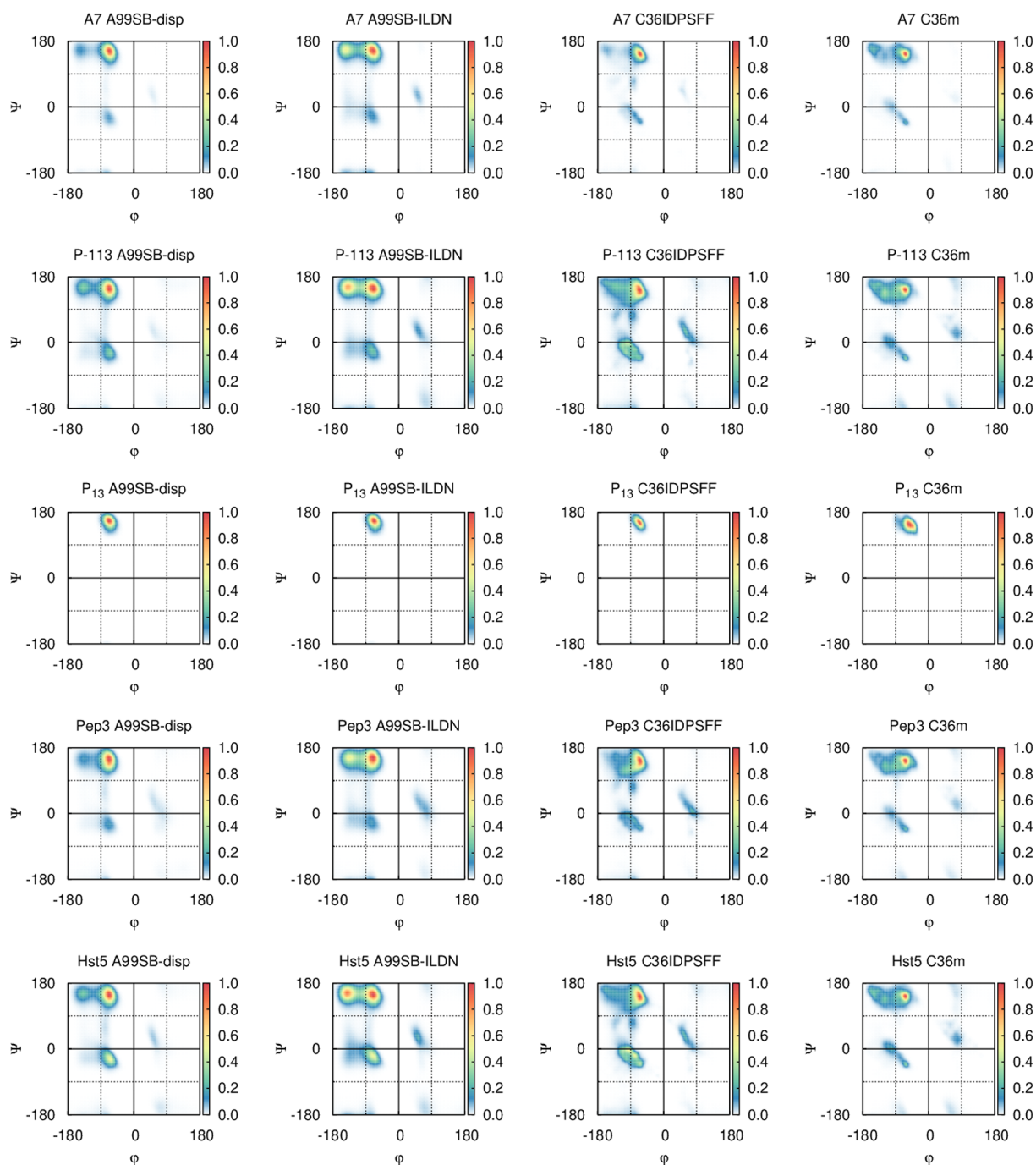
estimations from the Ramachandran plot. The last method is described in the following paragraph. The GROMACS tool `g_chi` was used to check the  $\omega$  angles in the simulations (see Figure S16).

The local structural preferences were also estimated from the dihedral angles of the peptide backbones, which were obtained by using the GROMACS tool `g_rama`. Only the  $\alpha$ -helical (both the right-handed and the left-handed), the  $\beta$ -sheet, and the PPII helical regions of the Ramachandran map were considered for this analysis. Similar to what has previously been done in many other studies,<sup>9,20,26,27,30,31,54–58</sup> a residue was considered to be in the right-handed  $\alpha$ -helix ( $\alpha_R$ ) region of the Ramachandran map when  $-90^\circ \leq \phi \leq -30^\circ$  and  $-90^\circ \leq \psi \leq 0^\circ$ , as illustrated in the Ramachandran map in Figure 1. Correspondingly,  $30^\circ \leq \phi \leq 90^\circ$  and  $0^\circ \leq \psi \leq 90^\circ$  were used for the left-handed  $\alpha$ -helix ( $\alpha_L$ ) region,  $-180^\circ \leq \phi < -104^\circ$  and  $180^\circ \leq \psi \leq 104^\circ$  were used for the  $\beta$ -sheet region, and  $-104^\circ \leq \phi \leq -46^\circ$  and  $116^\circ \leq \psi \leq 174^\circ$  were used for the PPII helix region. Residues not belonging to any of the aforementioned regions were unclassified but categorized as “bend/coil/turn” for simplification of the plots. We note that this classification is not based on secondary structure elements, but simply examines which regions of the Ramachandran map the different residues populate.

**2.2.3. CD Prediction.** To predict CD spectra from structural ensembles, we employed SESCA.<sup>59</sup> The SESCA algorithm has two steps:

1. The first step is per residue secondary structure assignment. We use DISICL<sup>60</sup> as the secondary structure prediction algorithm as it explicitly takes into account PPII conformations.

2. Spectral contributions from each secondary structure element in a conformation are combined to produce the CD spectra. In SESCA, the set of spectral contributions assigned to subsets of secondary structures are stored in the “basis sets”. Different basis sets for a given secondary structure assignment are available, which differ in the resolution of the spectral



**Figure 4.** Ramachandran plots aggregated over the full sequence of the five main peptides from simulations using four different force fields. The plots are normalized for a maximum intensity of 1.

contributions definition. Optionally, the spectral contributions of the side chains may be added. We tested several of the available basis sets, but mainly used DS6-1SC1 (DS6-1 with side chain contribution), as this gave rise to the predicted spectra that resembled the most in shape the experimental spectra.

Finally, the CD spectra from each conformation of the ensemble are linearly averaged.

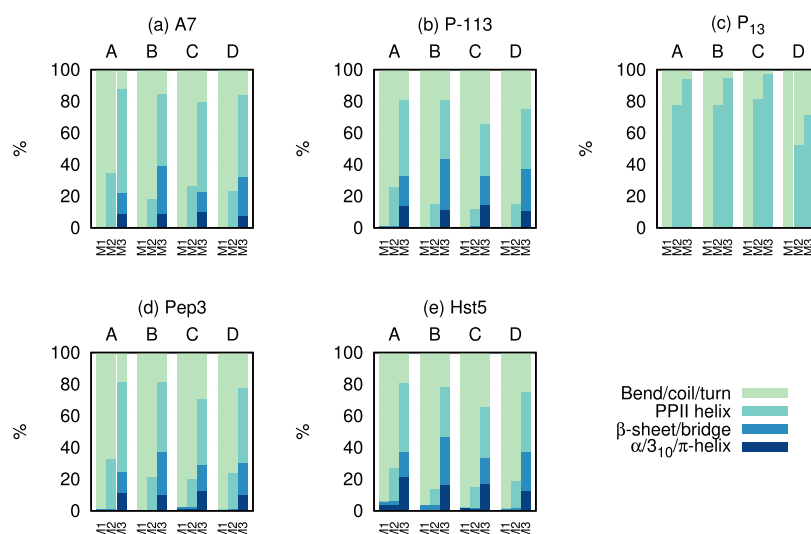
**2.3. Calculation of  $J$ -Couplings.** We used Karplus-like equations<sup>61</sup> to calculate the backbone NMR scalar ( $J$ ) couplings. This equation has the functional form

$$J(\theta) = A \cos^2(\theta + \delta) + B \cos(\theta + \delta) + C \quad (1)$$

In eq 1,  $\theta$  is the torsional angle that determines the  $J$ -coupling constant, while  $A$ ,  $B$ , and  $C$  are fitting parameters and  $\delta$  is a shift used in some calculations. For  ${}^3J_{\text{HNH}'}$ ,  ${}^3J_{\text{HC}'}$ ,  ${}^3J_{\text{HNC}'}$ , and  ${}^3J_{\text{HNC}''}$ , we used the parametrization from Lindorff-Larsen et al.<sup>62</sup> The same parametrizations employed in the work of Graf et al.<sup>63</sup> were used for  ${}^1J_{\text{NC}''}$ ,<sup>64</sup>  ${}^2J_{\text{NC}''}$ ,<sup>65</sup> and  ${}^3J_{\text{HNC}''}$ .<sup>66</sup>

To compare the calculated coupling constants with the experimental values we used

$$\chi^2 = \sum_{i=1}^N (\langle J_i \rangle_{\text{calc}} - J_{i,\text{exp}})^2 / \sigma_i^2 \quad (2)$$



**Figure 5.** Stacked histograms of average secondary structure of simulated peptides obtained from three different methods: M1 = DSSP, M2 = DSSP-PPII, and M3 = Ramachandran populations of the regions defined in Figure 1. We note that while M1 and M2 refer to assessments of secondary structure, M3 only reports on the sampling of regions of the Ramachandran map of each individual residue independent of the conformations of its neighbors.

where  $\langle J_i \rangle_{\text{calc}}$  is the time average of the  $i$ th  $J$ -coupling constant from the frames of a simulation,  $J_{i,\text{exp}}$  is the respective experimentally observed  $J$ -coupling constant, and  $\sigma_i$  is the error associated with the parametrization of the Karplus relationship.<sup>65–67</sup> Experimental errors and the errors on simulated averages are smaller than those from the Karplus parametrizations and were thus not included.

### 3. RESULTS AND DISCUSSION

**3.1. Effects of Force Field in Simulations of Five Peptides.** An initial comparison of the effects of the force fields was done by analyzing the resulting average radius of gyration of the five peptides (Table 4). All force fields resulted in similar average values for the radius of gyration for each of the individual peptides, although C36IDPSFF on average resulted in slightly more compact conformations compared to the other force fields. This was, however, not the case for  $P_{13}$ , for which C36m sampled a slightly smaller average instead. Overall, it seemed like both A99SB-disp and A99SB-ILDN sampled similar averages for all peptides.

Because the conformational ensembles of IDPs are highly heterogeneous, it is not trivial to find a set of variables that can describe the high variability of an ensemble in a low-dimensional representation. For each peptide, we used PCA (on aggregated trajectories over all force fields as discussed under Methods) to represent and visualize the simulations in a space of reduced dimensionality. After projecting the ensembles from the different force fields onto a common subspace, we examine the free energy surfaces projected as a function of the first two principal components, and in general we find relatively similar surfaces. However, the relative probabilities of the conformational states may differ, with C36IDPSFF giving rise to less “rough” surfaces, while the others show regions poorly explored at the simulated temperature (Figure 2).

Nonetheless, it is worth highlighting that, in the case of  $P_{13}$  with C36m, we observe a shift of the minimum on the second PC axis. Since  $P_{13}$  is thought to mostly populate PPII conformations, we decided to characterize and compare the free energy minima resulting from A99SB-disp and C36m.

Subtle differences were observed, both in the average radius of gyration (1.14 nm for A99SB-disp and 1.11 nm for CHARMM36m) and in the per residue average backbone dihedrals that, for both force fields, reside in the PPII ranges defined in DSSP-PPII<sup>20</sup> (Figure 3c). At the level of local structure, we find that A99SB-disp populates more PPII conformations than C36m (Figure 3d). Additionally, the PPII helix formed in simulations with A99SB-disp appears more bent with respect to an imaginary helix axis, while the PPII helix formed with C36m appears to be straighter (Figure 3a,b).

Ramachandran plots depicting all dihedral angles of each simulated peptide are presented in Figure 4. A few differences were observed when the force fields were compared. First, the Amber force fields (especially A99SB-ILDN) clearly show a more distinct  $\beta$ -sheet region and the populations in both the  $\alpha$ -helix region and the  $\beta$ -sheet region seem to be confined to smaller and more concentrated regions in the Amber simulations compared to the CHARMM simulations, where they seem to be spread out over larger areas. It is also worth noting that all Amber simulations have similar appearances/distributions over the Ramachandran space. The same is observed for the CHARMM simulations, and their appearance/distribution is slightly different from that of the Amber simulations. This strongly suggests that the different force field families (as might be expected) sample different structures. Interestingly, the PPII region seems to be somewhat shifted in the case of  $P_{13}$  with C36m, which is not seen for the other peptide simulations with the same force field. From studying these aggregated distributions across the Ramachandran plots alone, it is nearly impossible to obtain any detailed information on secondary structure propensities. Thus, the region populations have been quantified and are presented in Figure 5, where they are also compared to secondary structure estimates from the DSSP and the DSSP-PPII methods.

Comparisons of the average secondary structure content of each simulated peptide using three different methods are depicted in Figure 5. The DSSP analysis suggested that all peptides were fully disordered and dominated by bends, coils, and turns. Further investigation using DSSP-PPII did, however,



**Figure 6.** Stacked secondary structure histograms per amino acid residue of simulated peptides as obtained from DSSP-PPII algorithm.

reveal that approximately 10–35% of the secondary structure content was, in fact, PPII structure for most of the peptides. The  $P_{13}$  peptide was found to possess even more PPII structure (~50–80%), which is reasonable since it is expected to mainly possess PPII structure in aqueous solution.<sup>68,69</sup>

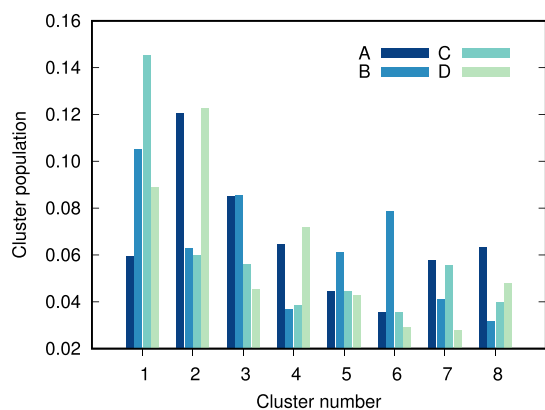
By comparing the secondary structure contents obtained from the different force fields using the DSSP-PPII method, A99SB-disp was found to sample more PPII structure than the other force fields for all peptides (except  $P_{13}$ ) and A99SB-disp and C36IDPSFF sampled slightly more  $\alpha/3_{10}/\pi$ -helical and  $\beta$ -sheet/bridge content. A7 and  $P_{13}$  possessed no other secondary structure elements according to this analysis. The amount of unstructured content (coil/bend/turn) was highest for C36m in

the case of  $P_{13}$ . No other obvious secondary structure propensities and trends were discerned.

Average populations of different regions of the Ramachandran map (corresponding to typical dihedral angles in different secondary structure elements) were also estimated from all of the simulations. Although this method was able to identify angles corresponding to PPII, it also showed large populations in the  $\alpha$ -helical and  $\beta$ -sheet regions, which were not observed in the DSSP and the DSSP-PPII analyses. This is not surprising since the Ramachandran map includes all angles regardless of position, whereas secondary structures need several consecutive amino acids with the same classification for them to register as a secondary structure. The Ramachandran analysis indicated that C36IDPSFF provided less sampling in the structured regions for



**Figure 7.** Representative structures of top eight cluster conformations of Hst5 as simulated with the four different force fields (A–D), as well as from a mixture of the four force fields (Mix). The total percentages of the top eight clusters are given above the structures, and the relative size of each individual cluster is given below each structure.



**Figure 8.** Weighted cluster population from the individual force fields in the top eight clusters of Hst5.

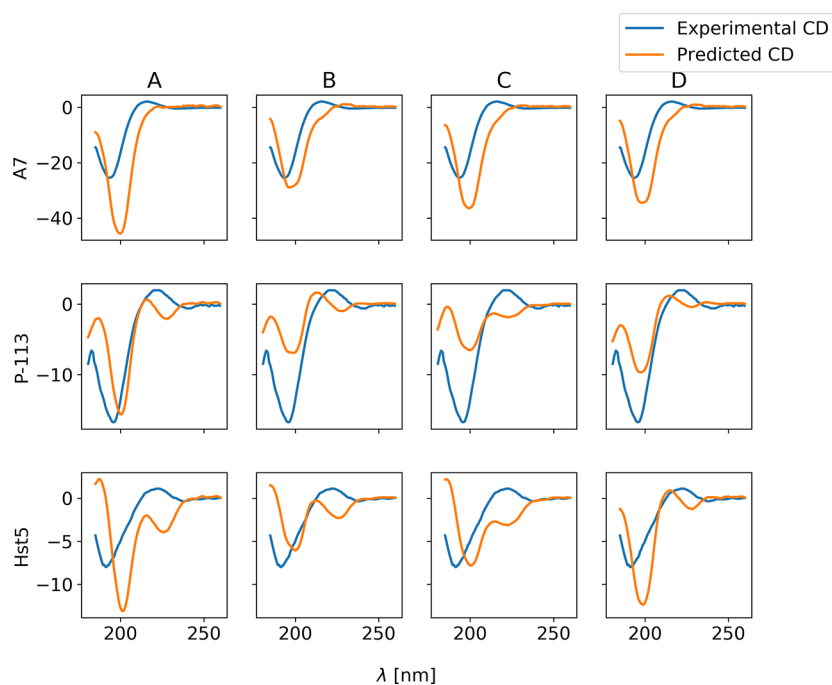
all peptides except P<sub>13</sub>, whereas A99SB-ILDN sampled larger populations in the  $\beta$ -sheet region than any other force field. Similar to what was observed in the DSSP-PPII analysis, A99SB-disp and C36IDPSFF were found to have larger sampling in the  $\alpha$ -helical regions than the other two force fields.

Each simulated peptide was investigated with DSSP-PPII to identify the average secondary structure content per amino acid residue (Figure 6). All force fields gave similar secondary structure profiles for A7 and P-113, although a small difference was observed between the Amber and the CHARMM force fields. In P-113 the largest PPII content was centered around

Arg-3 and Arg-9, and the lowest PPII content was found around Gly-6. The secondary structure profiles of P<sub>13</sub> were similar for all force fields except for C36IDPSFF, where the PPII content was lower. For the Pep3 simulations, the largest PPII content was centered around Ala-3 and Pro-7 for all force fields. A small increase in PPII content was also observed around Asn-11 in Pep3 for all force fields except C36IDPSFF. The PPII content followed a sharper decrease toward the C-terminus in the simulations with A99SB-ILDN and C36IDPSFF, and the lowest PPII content was centered around Gly-5. The  $\alpha/3_{10}/\pi$ -helical content in Pep3 was found mainly around Val-10 except for in the C36IDPSFF simulation, where a larger portion was centered around Lys-4. For Hst5, the largest PPII content was found around Lys-5 and Lys-13 for all force fields, in addition to a smaller peak around Lys-17. A99SB-ILDN and C36m had their  $\alpha/3_{10}/\pi$ -helical content around Tyr-10, whereas it was located closer to the termini in the C36IDPSFF simulations and around His-19 in the A99SB-disp simulations. The simulation of Hst5 with A99SB-disp also gave rise to a low amount of  $\alpha/3_{10}/\pi$ -helical content throughout most of the peptide. The  $\beta$ -sheet/bridge content in Hst5 was found around Lys-13 and Arg-22 in the Amber simulations but was more randomly distributed in the CHARMM simulations.

Differences between the force fields were further analyzed for Hst5. Cluster analysis was performed for each individual force field, as well as for a concatenated trajectory in which all four force fields were included. The analysis was done using a root-mean-square-deviation (RMSD) cutoff of 0.5 nm. This value





**Figure 9.** Comparison of experimental CD spectra with the ones predicted with SESCA from the conformational ensembles produced employing different force fields. The y-axes show the ellipticity,  $\theta$  (deg cm<sup>2</sup>/dmol).

**Table 5. Result of Comparison between the Experimental  $J$ -Coupling Constants of A7 and Those Calculated from the Conformational Ensembles Resulting from the Different Force Fields**

FF	$\chi^2$
A	19.8
B	98.1
C	39.6
D	53.1

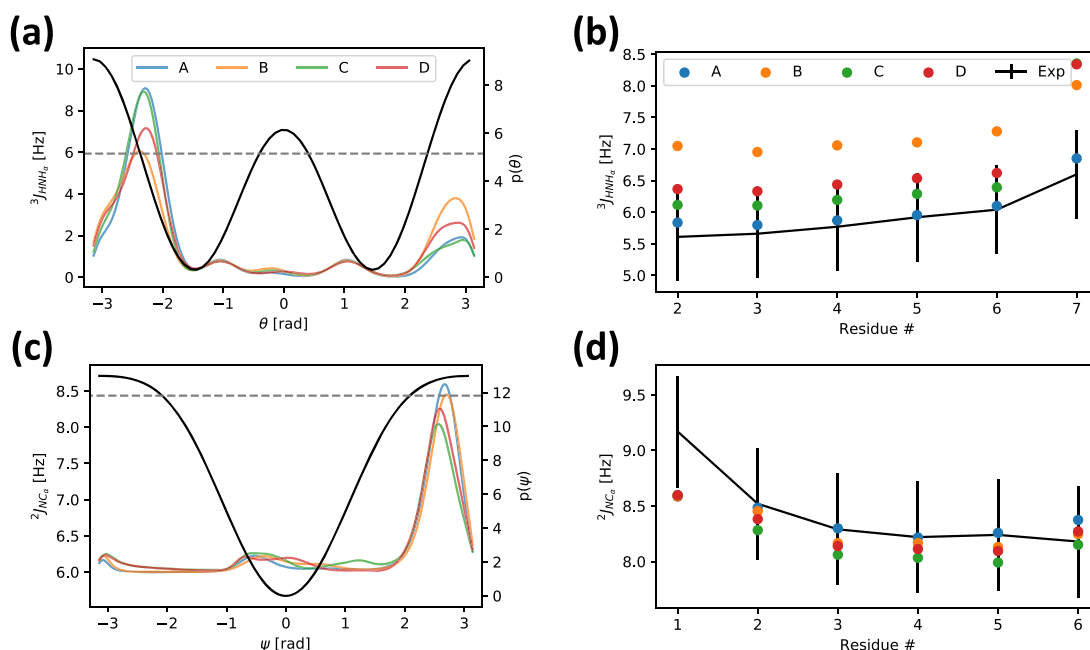
was chosen by examining the total population of the eight first clusters and varying the cluster radius (in steps of 0.05 nm) until their total population was closest to 50%. Additionally, using the same cutoff for all force fields made it easier to compare them. We do, however, note that because clustering methods compare “central structures”, they are not optimal to use for analyzing flexible peptides. Therefore, the number of clusters and their sizes that are presented here are not representations of the “truth”—they are simply used to compare the conformational sampling of the force fields.

The representative structures of the top eight conformational clusters of Hst5 with the four force fields, as well as the force field mix, are presented in Figure 7. Visual inspection of the representative structures immediately reveals that the first cluster conformers are different for the different force fields. Comparing the combined percentage sizes of the top eight conformation clusters gives some indication of the relative variability of the conformations sampled by the four different force fields. A higher value means that there are fewer conformations sampled in the remaining clusters, which suggests a lower degree of conformational variability. By this reasoning, of the four selected force fields, A99SB-disp provides the smallest amount of conformational variability with the remaining three force fields being comparable.

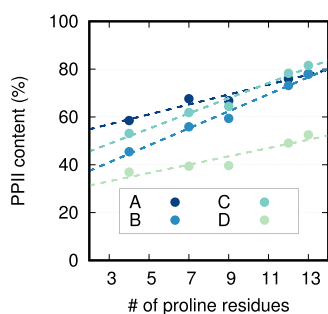
To get a more quantitative comparison among the force fields, their trajectories were concatenated, followed by a new cluster

analysis where each structure could be traced back to its individual force field. The relative cluster populations of the individual force fields in the top eight clusters are illustrated in Figure 8. Although all force fields are represented in each cluster, they are not evenly distributed. For example, the first cluster is dominated by C36IDPSFF, whereas the second cluster mainly contains conformations from A99SB-disp and C36m. The fifth cluster is the most evenly distributed cluster across the force fields, and the sixth cluster is heavily dominated by A99SB-ILDN. From this analysis it is safe to say that, although the average properties of different force fields may be similar, the force fields’ individual conformational ensembles are rather different, which naturally leads to different secondary structure contents.

**3.2. CD Prediction Using SESCA.** Since the four force fields give rise to different conformational ensembles, one may reasonably ask which of the force fields is more representative of the real conformational ensemble in solution. To answer this question, we may attempt to compare the simulations to experimental data. This ideally requires a forward model to predict the experimental observables from an ensemble of structures. We here used data from CD spectroscopy, as CD is highly sensitive of secondary structure composition, and we used SESCA<sup>59</sup> as a forward model. Experimental data for A7, P-113, and Hst5 were obtained from Graf et al.,<sup>63</sup> Han et al.,<sup>70</sup> and Jephthah et al.,<sup>48</sup> respectively. Unfortunately, and as also noted for other IDPs in the papers by Fagerberg et al.<sup>71</sup> and Gopal et al.,<sup>72</sup> it was not possible to obtain a meaningful agreement between the experimental CD spectra and those predicted by SESCA (Figure 9). This can be due to the fact that the main negative signature peak of a PPII helix may appear in experiments between 190 and 210 nm<sup>73–75</sup> because of non-secondary-structure contributions, while the spectral contribution associated with a PPII conformation in SESCA has a fixed position. Also, given the relative scarcity of PPII structure in folded proteins, it may be difficult to deconvolute its



**Figure 10.** Overview of two  $J$ -coupling constants predicted from A7 simulations. (a)  $^3J_{\text{HNH}_\alpha}$  and (c)  $^2J_{\text{NC}_\alpha}$  as a function of the underlying angles are represented as solid black lines (left y-axes), the experimental couplings (average over residues) are represented as dashed gray lines (left y-axes), and the dihedral angle distributions from the different force fields are represented in different colors (right y-axes). (b) and (d) show the experimental couplings (with errors used for the calculation of the  $\chi^2$  associated with the parametrizations of the Karplus relationship) as solid black lines, and the predicted couplings from the different force fields are represented with different colors.



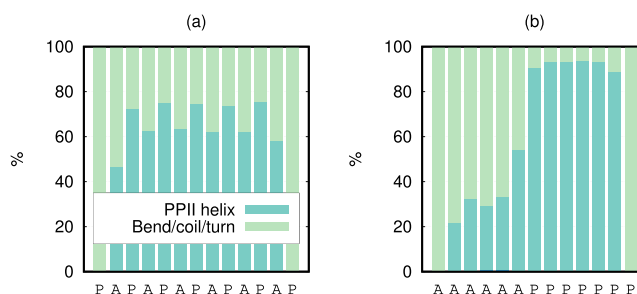
**Figure 11.** PPII helix content as obtained from DSSP-PPII analysis as a function of total number of Pro residues. Values from simulations of P<sub>13</sub>, V1, V2, V3, and V4, were used for this investigation.

**Table 6. Linear Regression Statistics of Figure 11<sup>a</sup>**

	$m$	$c$	$r^2$	$p$
A	2.076	50.691	0.948	0.005
B	3.561	30.322	0.983	0.001
C	3.181	39.218	0.973	0.002
D	1.738	27.875	0.870	0.021

<sup>a</sup>Slope ( $m$ ), intercept ( $c$ ), coefficient of determination ( $r^2$ ), and probability value ( $p$ ).

contribution when developing prediction methods for CD spectra. Also, a qualitative analysis based on the intensity of the main negative peak does not provide a decisive suggestion of what force field may be the most reliable. This is also complicated by some intensity scaling that may be needed to take into account uncertainty in the estimate of the concentration of the sample used for the experimental CD data. At this stage it is not clear if the source of the problem may be the force fields' inaccuracy, finite sampling, or the inaccuracy of the CD calculation for these kinds of systems.



**Figure 12.** Stacked secondary structure histograms per amino acid residue of (a) V2 and (b) V2.2 as obtained from the DSSP-PPII algorithm.

### 3.3. Comparing A7 Simulations to Experimental NMR

**Scalar Couplings.** A perhaps more accurate way of comparing simulations and experiments of flexible proteins and peptides is by investigating NMR scalar couplings. Since experimental  $J$ -coupling constants are available in the literature for A7<sup>63</sup> and  $J$ -couplings are more commonly used in comparison with computational models, we decided to calculate  $J$ -couplings from our A7 simulations (see [Methods](#)). Simulations with A99SB-disp show the best agreement with the experimental data, followed by C36IDPSFF, while C36m and A99SB-ILDN give rise to a less good fit with the experimental data ([Table 5](#)). This result may in part reflect that the target data for optimizing A99SB-disp included the experimental  $J$ -couplings for Ala<sup>58</sup>. With 39 scalar couplings used to calculate the  $\chi^2$  values, it appears that the deviations observed in C36m and A99SB-ILDN are greater than what would be expected by chance.

We examined in more detail the  $^3J_{\text{HNH}_\alpha}$  and  $^2J_{\text{NC}_\alpha}$  couplings, as  $^3J_{\text{HNH}_\alpha}$  can help discriminate the  $\phi$  dihedral regions of  $\beta$  and  $\alpha_1$ /PPII elements, and  $^2J_{\text{NC}_\alpha}$  can help discriminate the  $\psi$  dihedral

regions of  $\beta$ /PPII and  $\alpha_r$ .<sup>63</sup> We observe that the  $\psi$  angle distributions are relatively similar, with a strong peak corresponding to the  $\beta$ /PPII region, and the agreement with  $^2J_{NC_\alpha}$  is evenly good (Figure 10c,d). More differences can be observed for  $^3J_{HNH_r}$ , where especially A99SB-ILDN and C36m show a lower population in the  $\alpha_r$ /PPII (4–7 Hz<sup>76</sup>) and a higher population in an unclassified region, resulting in a lower agreement with the experimental  $^3J_{HNH_r}$  (Figure 10a,b).

**3.4. Effect of Proline Residue Content.** A few variants of P<sub>13</sub>, V1–V4 (see Table 1), were investigated to see how the Pro content affected the PPII propensities, as estimated by how much the residues sampled the PPII regions of the Ramachandran maps. Figure 11 shows the PPII content as a function of the number of Pro residues in P<sub>13</sub> and the peptide variants. All force fields yielded significant correlation ( $p < 0.05$ , see Table 6) between the PPII content and the number of Pro residues for P<sub>13</sub> and the chosen variants, where an increased number of Pro residues provided a larger PPII content. The slopes of these trends in the linear regression, however, differ depending on what force field was used, with C36m having the smallest increase and A99SB-ILDN having the largest increase. Furthermore, A99SB-ILDN was the force field that provided the strongest correlation (see Table 6).

V2.2 (see Table 1) was simulated using the A99SB-ILDN force field to investigate if the PPII content is affected by the relative position of the Pro residues in the amino acid sequence. The average PPII contents of V2 and V2.2 from the DSSP-PPII analysis were found to be essentially the same: 56 and 55%, respectively. Although the PPII content at first did not seem to be affected by the patterning of the Pro residues, histograms of the secondary structure per amino acid residue of V2 and V2.2 (Figure 12) revealed that the PPII content is significantly more localized to the Pro residues in V2.2, whereas it was more evenly distributed in V2. However, further investigation is needed to fully characterize this trend. It is, for example, necessary to study the effect of patterning of Pro in the other P<sub>13</sub> variants as well. It would also be of interest to see how the trend is affected by the peptide length.

## 4. CONCLUSIONS

In this study we have evaluated the differences among four different force fields in simulations of five short peptides with varying PPII propensities. All force fields gave similar ensemble averages of the radius of gyration, although the averages by the C36IDPSFF force field were generally slightly smaller compared to the other force fields. All force fields appeared to sample comparable regions of conformational spaces (as probed by the first two principal components) for each individual peptide, although with slightly different probabilities. Similarly, all force fields sampled the PPII structure, but to different extents. Additionally, some force fields were more prone to sampling other secondary structure elements. For example, A99SB-disp and C36IDPSFF sampled more  $\alpha/3_{10}/\pi$ -helical and  $\beta$ -sheet/bridge content, C36m sampled less structured content than the other force fields, and A99SB-disp often had the highest PPII content. Direct comparison by conformational clustering revealed that the force fields have a bias toward different conformational clusters. CD prediction using SESCA was performed to examine which force field provided a more accurate conformational ensemble. Unfortunately, the method was not able to match the predicted and experimental spectra. We also calculated scalar couplings and compared them to

experimental results for A7 and found two force fields (A99SB-disp and C36IDPSFF) that gave agreement roughly within expected error. We note that the calculations of both CD and scalar couplings contain contributions from all types of local structures and do not solely report on the accuracy of the PPII content. Finally, we investigated the effect of Pro residue content on the PPII content of short peptides containing only Ala and Pro, and we observed a correlation between the number of Pro residues in the amino acid sequence and the PPII content.

We conclude by highlighting that we need better methods to calculate experimental observables that are sensitive to secondary structure preferences for flexible peptides. Such methods are often parametrized using folded protein structures and, thus, may be difficult to apply or inaccurate for disordered peptides and proteins.<sup>77</sup> In particular, we stress the need for better methods to link populations of PPII-like structures in simulations to a broader range of experimental observables and note that NMR chemical shifts can also be used for this purpose.<sup>78,79</sup> This is especially needed for simulations of proteins in which PPII might have a significant role, such as for example Hst5, SH3-binding peptides, and collagen.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00408>.

Probability distributions, autocorrelation functions, block average error estimates, and  $\omega$  angle as a function of time (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Kresten Lindorff-Larsen – Structural Biology and NMR Laboratory & the Linderstrom-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark; [orcid.org/0000-0002-4750-6039](https://orcid.org/0000-0002-4750-6039); Phone: +45 35 32 20 27; Email: [lindorff@bio.ku.dk](mailto:lindorff@bio.ku.dk)

Marie Skepö – Division of Theoretical Chemistry, Lund University, SE-221 00 Lund, Sweden; [orcid.org/0000-0002-8639-9993](https://orcid.org/0000-0002-8639-9993); Phone: +46 46 222 33 66; Email: [marie.skepo@teokem.lu.se](mailto:marie.skepo@teokem.lu.se)

### Authors

Stéphanie Jephthah – Division of Theoretical Chemistry, Lund University, SE-221 00 Lund, Sweden; [orcid.org/0000-0003-4287-506X](https://orcid.org/0000-0003-4287-506X)

Francesco Pesce – Structural Biology and NMR Laboratory & the Linderstrom-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00408>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was funded by the NordForsk's Nordic Neutron Science Programme (Grant No. 81912) and the Lundbeck Foundation BRAINSTRUC initiative in structural biology (R155-2015-2666, [lundbeckfonden.com](http://lundbeckfonden.com)). Computer resources

for the simulations were provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC). We also acknowledge access to computational resources from the ROBUST Resource for Biomolecular Simulations (supported by the Novo Nordisk Foundation Grant No. NF18OC0032608) and from the Biocomputing Core Facility at the Department of Biology, University of Copenhagen.

## REFERENCES

- (1) Bernadó, P.; Svergun, D. I. Structural insights into intrinsically disordered proteins by small-angle X-ray scattering. Instrumental analysis of intrinsically disordered proteins. *Assessing structure and conformation* **2010**, 451–476.
- (2) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J.* **2008**, *95*, L07.
- (3) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.
- (4) Henriques, J.; Cragnell, C.; Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420–3431.
- (5) Krimm, S.; Tiffany, M. L. The circular dichroism spectrum and structure of unordered polypeptides and proteins. *Isr. J. Chem.* **1974**, *12*, 189–200.
- (6) Adzhubei, A. A.; Sternberg, M. J.; Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **2013**, *425*, 2100–2132.
- (7) Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958.
- (8) Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S. A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *J. Comput. Chem.* **2005**, *26*, 1612–1616.
- (9) Aliev, A. E.; Courtier-Murias, D. Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *J. Phys. Chem. B* **2010**, *114*, 12358–12375.
- (10) Musacchio, A.; Wilmanns, M.; Saraster, M. Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* **1994**, *61*, 283–297.
- (11) Rath, A.; Davidson, A. R.; Deber, C. M. The structure of “unstructured” regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers* **2005**, *80*, 179–185.
- (12) Kaneko, T.; Li, L.; Li, S. The SH3 domain—a family of versatile peptide-and protein-recognition module. *Front. Biosci., Landmark Ed.* **2008**, *13*, 4938–4952.
- (13) Kakinoki, S.; Hirano, Y.; Oka, M. On the stability of polyproline-I and II structures of proline oligopeptides. *Polym. Bull.* **2005**, *53*, 109–115.
- (14) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. Conformation of the backbone in unfolded proteins. *Chem. Rev.* **2006**, *106*, 1877–1897.
- (15) Zanna, N.; Milli, L.; Del Secco, B.; Tomasini, C. Factors affecting the stabilization of polyproline II helices in a hydrophobic environment. *Org. Lett.* **2016**, *18*, 1662–1665.
- (16) Uversky, V. N.; Dunker, A. K. *Intrinsically Disordered Protein Analysis: Vol. 1, Methods and Experimental Tools*; Springer: 2012.
- (17) Provencher, S. W.; Gloeckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **1981**, *20*, 33–37.
- (18) Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. 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.
- (19) Whitmore, L.; Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **2008**, *89*, 392–400.
- (20) Mansiaux, Y.; Joseph, A. P.; Gelly, J.-C.; de Brevern, A. G. Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PLoS One* **2011**, *6*, No. e18401.
- (21) Chebrek, R.; Leonard, S.; de Brevern, A. G.; Gelly, J.-C. PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* **2014**, *2014*, bau102.
- (22) Chong, S.-H.; Chatterjee, P.; Ham, S. Computer simulations of intrinsically disordered proteins. *Annu. Rev. Phys. Chem.* **2017**, *68*, 117–134.
- (23) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154.
- (24) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129–138.
- (25) Zerze, G. H.; Zheng, W.; Best, R. B.; Mittal, J. Evolution of All-atom Protein Force Fields to Improve Local and Global Properties. *J. Phys. Chem. Lett.* **2019**, *10*, 2227–2234.
- (26) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (27) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.
- (28) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950–1958.
- (29) Jiang, F.; Zhou, C.-Y.; Wu, Y.-D. Residue-specific force field based on the protein coil library. RSFF1: modification of OPLS-AA/L. *J. Phys. Chem. B* **2014**, *118*, 6983–6998.
- (30) 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.
- (31) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.
- (32) Song, D.; Wang, W.; Ye, W.; Ji, D.; Luo, R.; Chen, H.-F. ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. *Chem. Biol. Drug Des.* **2017**, *89*, 5–15.
- (33) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (34) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (35) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (36) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (37) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (38) 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.
- (39) Liu, H.; Song, D.; Lu, H.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Chem. Biol. Drug Des.* **2018**, *92*, 1722–1735.

- (40) Liu, H.; Song, D.; Zhang, Y.; Yang, S.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21918–21931.
- (41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (42) *The PyMOL Molecular Graphics System*, ver. 1.2r1; Schrödinger LLC: 2009.
- (43) Berendsen, H. J. C.; Van Gunsteren, W. F. Practical algorithms for dynamic simulations. In *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*; North-Holland: 1986; pp 43–65.
- (44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.
- (45) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (46) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (47) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182.
- (48) Jephthah, S.; Staby, L.; Kragelund, B.; Skepo, M. Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? *J. Chem. Theory Comput.* **2019**, *15*, 2672–2683.
- (49) Hess, B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *J. Chem. Phys.* **2002**, *116*, 209–217.
- (50) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
- (51) Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing molecular dynamics force fields in the essential subspace. *PLoS One* **2015**, *10*, No. e0121114.
- (52) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.
- (53) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- (54) García, A. E.; Sanbonmatsu, K. Y.  $\alpha$ -Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2782–2787.
- (55) García, A. E. Characterization of non-alpha helical conformations in Ala peptides. *Polymer* **2004**, *45*, 669–676.
- (56) Song, K.; Stewart, J. M.; Fesinmeyer, R. M.; Andersen, N. H.; Simmerling, C. Structural insights for designed alanine-rich helices: Comparing NMR helicity measures and conformational ensembles from molecular dynamics simulation. *Biopolymers* **2008**, *89*, 747–760.
- (57) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (58) Yu, L.; Li, D.-W.; Brüschweiler, R. Systematic Differences between Current Molecular Dynamics Force Fields To Represent Local Properties of Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2021**, *125*, 798–804.
- (59) Nagy, G.; Igaev, M.; Jones, N. C.; Hoffmann, S. V.; Grubmüller, H. SESCA: predicting circular dichroism spectra from protein molecular structures. *J. Chem. Theory Comput.* **2019**, *15*, 5087–5102.
- (60) Nagy, G.; Oostenbrink, C. Dihedral-based segment identification and classification of biopolymers I: proteins. *J. Chem. Inf. Model.* **2014**, *54*, 266–277.
- (61) Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11–15.
- (62) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. Interpreting Dynamically-Averaged Scalar Couplings in Proteins. *J. Biomol. NMR* **2005**, *32*, 273–280.
- (63) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
- (64) Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2002**, *23*, 47–55.
- (65) Ding, K.; Gronenborn, A. M. Protein Backbone  $^1\text{H-N-}^{13}\text{C}\alpha$  and  $^{15}\text{N-}^{13}\text{C}\alpha$  Residual Dipolar and J Couplings: New Constraints for NMR Structure Determination. *J. Am. Chem. Soc.* **2004**, *126*, 6232–6233.
- (66) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. Determination of  $\psi$  Torsion Angle Restraints from  $^3\text{J}(\text{C}\alpha, \text{C}\alpha)$  and  $^3\text{J}(\text{C}\alpha, \text{HN})$  Coupling Constants in Proteins. *J. Am. Chem. Soc.* **2000**, *122*, 6268–6277.
- (67) Hu, J.-S.; Bax, A. Determination of  $\phi$  and  $\chi^1$  Angles in Proteins from  $^{13}\text{C-}^{13}\text{C}$  Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc.* **1997**, *119*, 6360–6368.
- (68) Tiffany, M. L.; Krimm, S. Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers* **1968**, *6*, 1767–1770.
- (69) Dukor, R. K.; Keiderling, T. A. Mutarotation studies of poly-L-proline using FTIR, electronic and vibrational circular dichroism. *Biospectroscopy* **1996**, *2*, 83–100.
- (70) Han, J.; Jyoti, M. A.; Song, H.-Y.; Jang, W. S. Antifungal activity and action mechanism of histatin 5-halocidin hybrid peptides against *Candida* ssp. *PLoS One* **2016**, *11*, No. e0150196.
- (71) Fagerberg, E.; Månsson, L. K.; Lenton, S.; Skepö, M. The Effects of Chain Length on the Structural Properties of Intrinsically Disordered Proteins in Concentrated Solutions. *J. Phys. Chem. B* **2020**, *124*, 11843–11853.
- (72) Gopal, S. M.; Wingbermühle, S.; Schnatwinkel, J.; Juber, S.; Herrmann, C.; Schäfer, L. V. Conformational Preferences of an Intrinsically Disordered Protein Domain: A Case Study for Modern Force Fields. *J. Phys. Chem. B* **2021**, *125*, 24–35.
- (73) Lopes, J. L.; 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.
- (74) Rucker, A. L.; Creamer, T. P. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Science* **2002**, *11*, 980–985.
- (75) Woody, R. W. Circular dichroism spectrum of peptides in the poly (Pro) II conformation. *J. Am. Chem. Soc.* **2009**, *131*, 8234–8245.
- (76) Mueller, G. A.; Smith, A. M.; Williams, D. C.; Hakkaart, G. A.; Aalberse, R. C.; Chapman, M. D.; Rule, G. S.; Benjamin, D. C. Expression and Secondary Structure Determination by NMR Methods of the Major House Dust Mite Allergen Der p 2. *J. Biol. Chem.* **1997**, *272*, 26893–26898.
- (77) Lindorff-Larsen, K.; Kragelund, B. B. On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *J. Mol. Biol.* **2021**, 167196.
- (78) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **2012**, *51*, 2224–2231.
- (79) Krieger, J. M.; Fusco, G.; Lewitzky, M.; Simister, P. C.; Marchant, J.; Camilloni, C.; Feller, S. M.; De Simone, A. Conformational recognition of an intrinsically disordered protein. *Biophys. J.* **2014**, *106*, 1771–1779.