



Article

# Molecular Dynamics Simulations of Phosphorylated Intrinsically Disordered Proteins: A Force Field Comparison

Ellen Rieloff <sup>1</sup> and Marie Skepö <sup>1,2,\*</sup>

<sup>1</sup> Division of Theoretical Chemistry, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden; ellen.rieloff@teokem.lu.se

<sup>2</sup> LINXS—Lund Institute of Advanced Neutron and X-ray Science, Scheelevägen 19, SE-223 70 Lund, Sweden

\* Correspondence: marie.skepö@teokem.lu.se

**Abstract:** Phosphorylation is a common post-translational modification among intrinsically disordered proteins and regions, which helps regulate function by changing the protein conformations, dynamics, and interactions with binding partners. To fully comprehend the effects of phosphorylation, computer simulations are a helpful tool, although they are dependent on the accuracy of the force field used. Here, we compared the conformational ensembles produced by Amber ff99SB-ILDN+TIP4P-D and CHARMM36m, for four phosphorylated disordered peptides ranging in length from 14–43 residues. CHARMM36m consistently produced more compact conformations with a higher content of bends, mainly due to more stable salt bridges. Based on comparisons with experimental size estimates for the shortest and longest peptide, CHARMM36m appeared to overestimate the compactness. The difference between the force fields was largest for the peptide showing the greatest separation between positively charged and phosphorylated residues, in line with the importance of charge distribution. For this peptide, the conformational ensemble did not change significantly upon increasing the ionic strength from 0 mM to 150 mM, despite a reduction of the salt-bridging probability in the CHARMM36m simulations, implying that salt concentration has negligible effects in this study.

**Keywords:** intrinsically disordered proteins; phosphorylation; force fields



**Citation:** Rieloff, E.; Skepö, M. Molecular Dynamics Simulations of Phosphorylated Intrinsically Disordered Proteins: A Force Field Comparison. *Int. J. Mol. Sci.* **2021**, *22*, 10174. <https://doi.org/10.3390/ijms221810174>

Academic Editors: Istvan Simon and Csaba Magyar

Received: 1 September 2021

Accepted: 16 September 2021

Published: 21 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intrinsically disordered proteins (IDPs) are characterized by a lack of a tertiary structure under physiological conditions [1,2], which means that they are better described by an ensemble of different conformations than a single structure. This is reflected in their free energy landscapes, which normally are rather flat without a deep energy minimum as for globular proteins [3]. The flattened energy landscape makes IDPs very sensitive to changes in the environment and post-translational modifications (PTMs) of the sequence. A common type of reversible PTM is phosphorylation, which introduces extra negative charges and the possibility of forming hydrogen bonds and salt bridges [4]. Phosphorylation is commonly employed by cells as a regulatory mechanism, as it can change both the conformational ensemble and the dynamics, as well as the interaction with a binding partner, and therefore affect function. The functional implications of phosphorylation can be drastic, such as for the disordered neuroprotein tau, for which hyperphosphorylation has been related to amyloid fibril formation in Alzheimer's disease [5]. In proteins such as statherin and caseins, the phosphorylated residues are essential for their ability to bind to the tooth surface [6,7] or sequester calcium [8].

Experimental techniques such as small-angle X-ray scattering (SAXS) and fluorescence resonance energy transfer (FRET) have been used to provide information on global conformational changes upon phosphorylation of intrinsically disordered proteins or regions, while circular dichroism spectroscopy and nuclear magnetic resonance (NMR) have detected changes in secondary structure or other local arrangements such as salt

bridges [9–14]. However, due to the vast conformational ensembles possessed by IDPs, computer simulations are often a useful complement to obtain more detailed information, though this requires accurate models and force fields. We have previously shown that a coarse-grained “one bead per residue model” has proven to accurately predict average radius of gyration ( $R_g$ ) and scattering curves for various IDPs, including statherin, although producing overly compact conformations of other more phosphorylated IDPs [15]. The two-site UNRES model has recently been extended with parameters for phosphorylated residues [16] and applied to study phosphorylation-induced folding of an IDP [17]. Although coarse-grained models are more computationally efficient and generally easier to interpret than atomistic models, they can lack in detail. In atomistic modelling, there is continuous development of force fields and water models towards more accurately describing IDPs, and some important adjustments have been the refinement of the backbone dihedral angles and balancing the water–protein and protein–protein interactions; see for example the following reviews and references within [18,19]. However, we recently showed that while the commonly used force fields CHARMM36m and Amber ff99SB-ILDN+TIP4P-D accurately captured the global dimensions of the 15-residue-long N-terminal fragment of Statherin in the nonphosphorylated state, it overestimated the compactness in the phosphorylated state [20]. More recently, overcompaction was also observed for two approximately 80-residue-long phosphorylated IDPs in several force fields, where it was suggested to depend on an overestimation of charge–charge interactions [21], in line with an over-stabilization of salt bridges in standard force fields [22]. In this study, we made a further comparison of the two aforementioned force fields, by applying them to four phosphorylated peptides, namely two different fragments from tau, specifically residues 173–183 (Tau1) and 225–246 (Tau2), the first 25 amino acids in the milk protein  $\beta$ -casein (bCPP) and the saliva protein statherin (Stath). For all peptides, CHARMM36m was shown to sample more compact conformations than Amber ff99SB-ILDN+TIP4P-D, associated with a much higher probability for salt bridges. The effect was more pronounced in sequences with large separation between phosphorylated residues and positively charged residues, showing the importance of charge distribution. In bCPP, which showed the largest differences between the force fields, the addition of 150 mM NaCl did not change the average size estimates and shape significantly, despite a significant reduction of salt bridge occurrence in CHARMM36m. This implies that salt bridges are still of importance at 150 mM salt and that we can ignore the effects of salt concentration in this study.

## 2. Results and Discussion

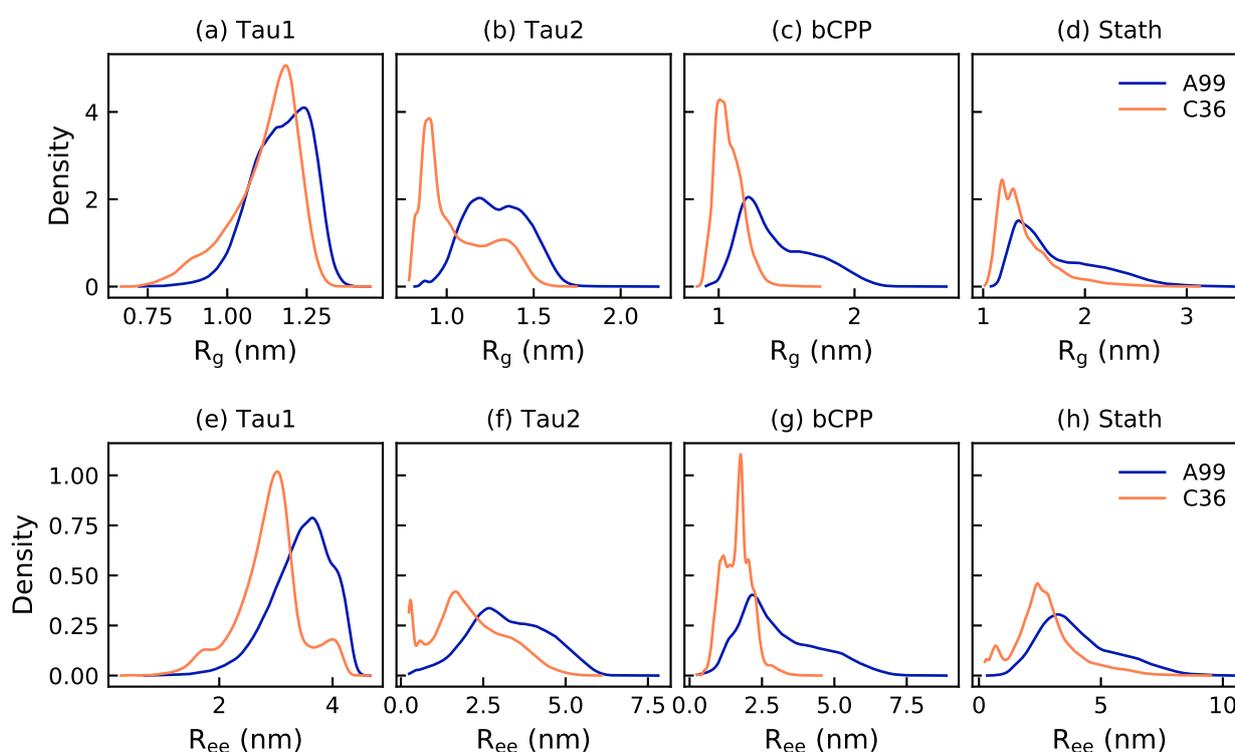
Four phosphorylated peptides, shown in Table 1, were simulated at physiological pH using two different force fields: Amber ff99SB-ILDN [23] with the TIP4P-D [24] water model and parameters for the phosphorylated residues from Homeyer et al. [25] and Steinbrecher et al. [26] (A99) and CHARMM36m [27] with the CHARMM-modified TIP3P water model [28] (C36). The peptides were chosen based on availability of experimental data to compare with and size considering the computational expense.

**Table 1.** Full name and sequence of the peptides included in this study. Positively charged residues are marked in blue, negatively charged in red, and phosphorylated residues highlighted with yellow. Note that Tau1 includes three additional residues in accordance with [11], to allow for experimental comparison.

Name	Protein	Sequence
Tau1	Tau <sub>173-183</sub>	CAK <sup>+</sup> PPAPK <sup>+</sup> PPAW
Tau2	Tau <sub>225-246</sub>	KVAVVR <sup>+</sup> PPK <sup>+</sup> SPSSAKSRLQTA
bCPP	$\beta$ -casein <sub>1-25</sub>	RELEELNVPGEIVE <sup>-</sup> LS <sup>-</sup> SSSEESITR
Stath	Statherin	DS <sup>-</sup> SEEKFLRRIGR <sup>+</sup> FGYGYGPYQPVPEQPLYPQPYQPQYQQYTF

### 2.1. Size and Shape

For all four peptides, the two force fields produced different conformational ensembles, as seen by the distributions of the  $R_g$  and the end-to-end distance ( $R_{ee}$ ) in Figure 1. The C36 distributions were narrower and centered on values lower than the A99 distributions. For Tau2 and bCPP, the  $R_g$  distribution had a sharp peak at low values. From the average  $R_g$  and  $R_{ee}$  presented in Table 2, it is clear that Tau1 showed the smallest differences between the force fields, while bCPP showed the largest differences. The discrepancy was larger for  $R_{ee}$  than  $R_g$ . For Tau1, Chin et al. [11] determined the average  $R_{ee}$  to be  $\sim 3.17$  nm, based on FRET. To obtain an  $R_{ee}$  distance distribution from the FRET data they assumed a semi-flexible polymer model, and the resulting distribution was skewed towards longer distances, with the peak value located at 3.64 nm (Figure 4A in ref. [11]). Comparing A99 and C36 to the experimental average, A99 overestimated it approximately as much as C36 underestimated it. However, the skewed shape and peak position at 3.64 nm produced in A99 was in better experimental agreement than C36, since the distribution in C36 was more symmetrical with multiple peaks and had the main peak located at 3.03 nm.



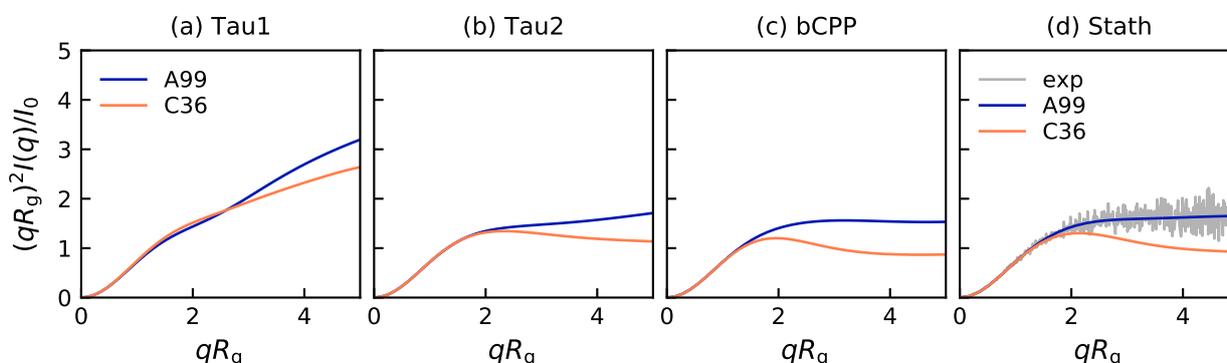
**Figure 1.** Distribution of the radius of gyration (**top row**) and the end-to-end distance (**bottom row**) of Tau1, Tau2, bCPP, and Stath simulated with Amber ff99SB-ILDN (A99) and CHARMM36m (C36). The legend applies to all panels.

**Table 2.** Average radius of gyration and end-to-end distance of the peptides simulated with Amber ff99-SB-ILDN (A99) and CHARMM36m (C36). The difference between the force fields is expressed in relation to A99.

Peptide	Radius of Gyration (nm)			End-to-End Distance (nm)		
	A99	C36	Difference (%)	A99	C36	Difference (%)
Tau1	$1.17 \pm 0.01$	$1.12 \pm 0.01$	4	$3.44 \pm 0.04$	$2.88 \pm 0.07$	16
Tau2	$1.29 \pm 0.03$	$1.06 \pm 0.10$	18	$3.27 \pm 0.17$	$2.10 \pm 0.32$	36
bCPP	$1.43 \pm 0.03$	$1.08 \pm 0.02$	24	$3.09 \pm 0.15$	$1.65 \pm 0.10$	47
Stath	$1.73 \pm 0.09$	$1.41 \pm 0.04$	18	$4.05 \pm 0.17$	$2.74 \pm 0.20$	32

For Stath, earlier published SAXS data [15] provided an  $R_g$  of  $1.93 \pm 0.2$  nm; hence,  $R_g$  was 10% smaller in A99 and 27% smaller in C36. Since  $R_g$  determined from SAXS includes a hydration shell, it was expected that  $R_g$  calculated from simulations would be slightly smaller, although not to that extent. Since it is not straightforward which contrast to use for the hydration shell in the calculations of scattering curves for IDPs [29], in Supplementary Figure S1 and Table S2, we compared the curves calculated using different contrasts of the hydration shell to the experimental curve for Stath. While the highest contrast used ( $0.03 e/\text{\AA}^3$ ) yielded the best agreement with the scattering curve, it provided the worst agreement with the Kratky plot. Henriques et al. [29] showed that the optimal contrast for IDPs was often between  $0.01 e/\text{\AA}^3$  and  $0.02 e/\text{\AA}^3$ , although varying with both force field and protein. The optimal values for A99 and C36 were suggested to be around  $0.0075 e/\text{\AA}^3$  and  $0.02 e/\text{\AA}^3$ , respectively. While the suggested optimal value gave reasonable agreement with the experimental form factor for A99, this was not the case for C36. For C36, all contrasts  $> 0$  clearly showed larger compaction than the experimental Kratky plot.

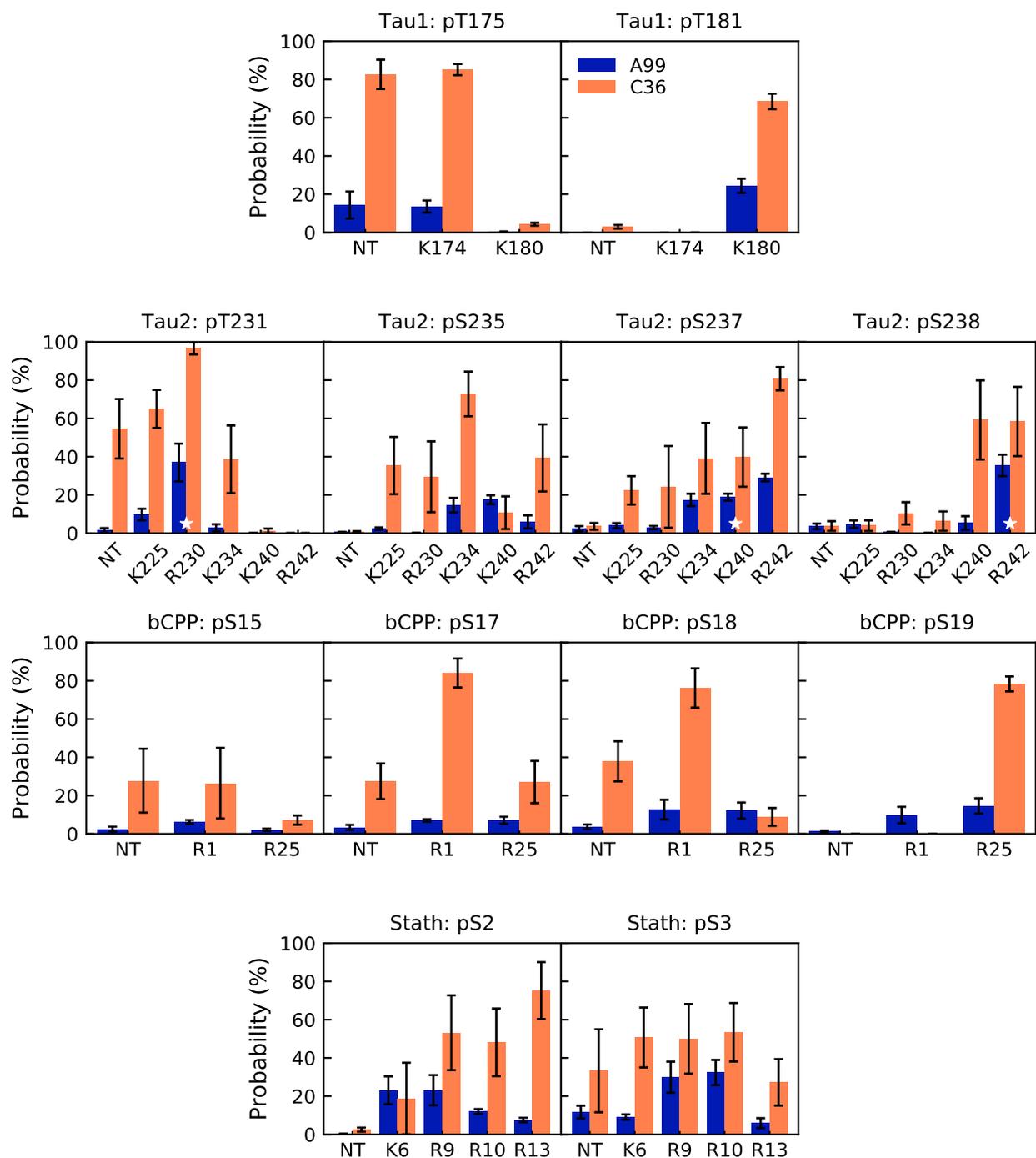
Even without experimental scattering curves to compare to, the dimensionless Kratky plot, presented in Figure 2, is a good way of comparing the average shape of the peptides in the two different force fields. The short peptide Tau1 exhibited a more extended shape than the other three peptides, which in A99 were shown to have more of the typical IDP behavior, resembling a Gaussian chain. For all four peptides, the Kratky plot produced in C36 had a lower slope, and for the three longest peptides, the curve started to move towards the bell-shaped curve typical of globular proteins. Hence, this implies that C36 sampled more compact or well-defined conformations than A99, in accordance with the  $R_g$  and  $R_{ee}$  distributions. Notice also that the Kratky plot of Stath in A99 was in excellent agreement with the experimental data, while the curve corresponding to C36 fell below, as shown in Figure 2d.



**Figure 2.** Dimensionless Kratky plot from simulations with Amber ff99SB-ILDN and CHARMM36m for (a) Tau1, (b) Tau2, (c) bCPP, and (d) Stath. In Panel (d), experimental data from Craggell et al. [15] are included for comparison. The legend in Panel (a) is applicable to all panels.

## 2.2. Salt Bridges and Secondary Structure

Since our previous study [20] suggested that overstabilized salt bridges are the reason why C36 produces more compact conformations than A99, we calculated the occupancy of the possible salt bridge interactions involving the phosphorylated residues. Figure 3 indeed shows that salt bridges were formed much more in C36 than A99, for all the peptides. In Tau2 and bCPP, the strong salt bridges in C36 restricted the conformational ensemble, which explains the smaller and narrower distributions of  $R_g$  and  $R_{ee}$ . In bCPP, the salt-bridging residues were well separated in the sequence, therefore having a larger effect on the  $R_g$  and  $R_{ee}$  distributions. In Tau1, the salt bridge interactions almost exclusively appeared between the adjacent residues and between pT175 and the N-terminal.

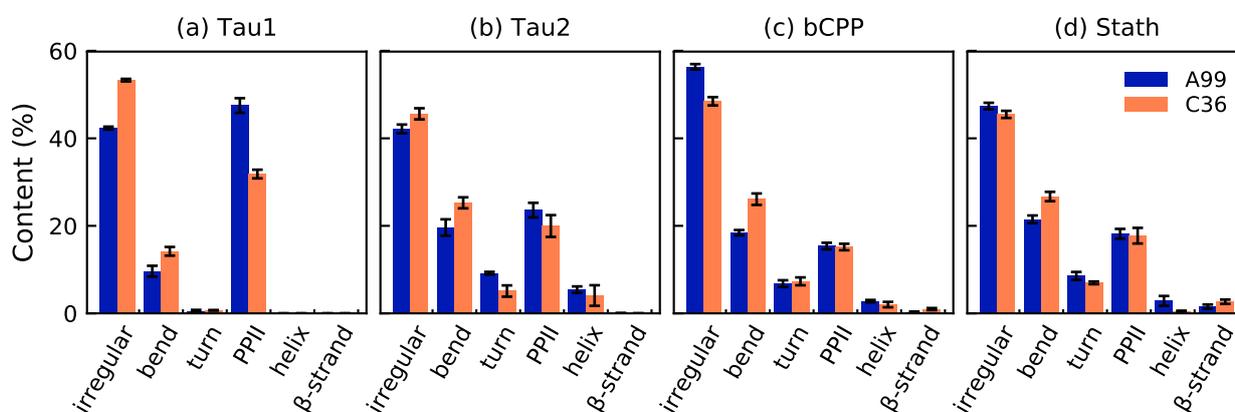


**Figure 3.** Probability of possible salt bridge interactions for the phosphorylated residues with the N-terminus (NT) and positively charged residues in Tau1 (first row), Tau2 (second row), bCPP (third row), and Stath (last row). For Tau2, experimentally established salt bridges [12] are marked with a white star. Error bars correspond to errors calculated by block averaging.

For Tau2, there is experimental evidence of the following salt bridges, detected by NMR experiments: pT231–R230, pS237–K240, and pS238–R242 [12]. pT231–R230 and pS238–R242 are indeed two of the most often occurring salt bridges in A99, while pS237–R242 is more common than pS237–K240. Several other salt bridges are also as frequently present as pS237–K240. In C36, pT231–R230 is the most occurring salt bridge, but both pS327–R242 and pS235–K234 are more probable than pS237–K240. Hence, while both

force fields captured the experimentally established salt bridges, they also suggested other salt bridges to be present and some of them to be more common than the experimentally established ones.

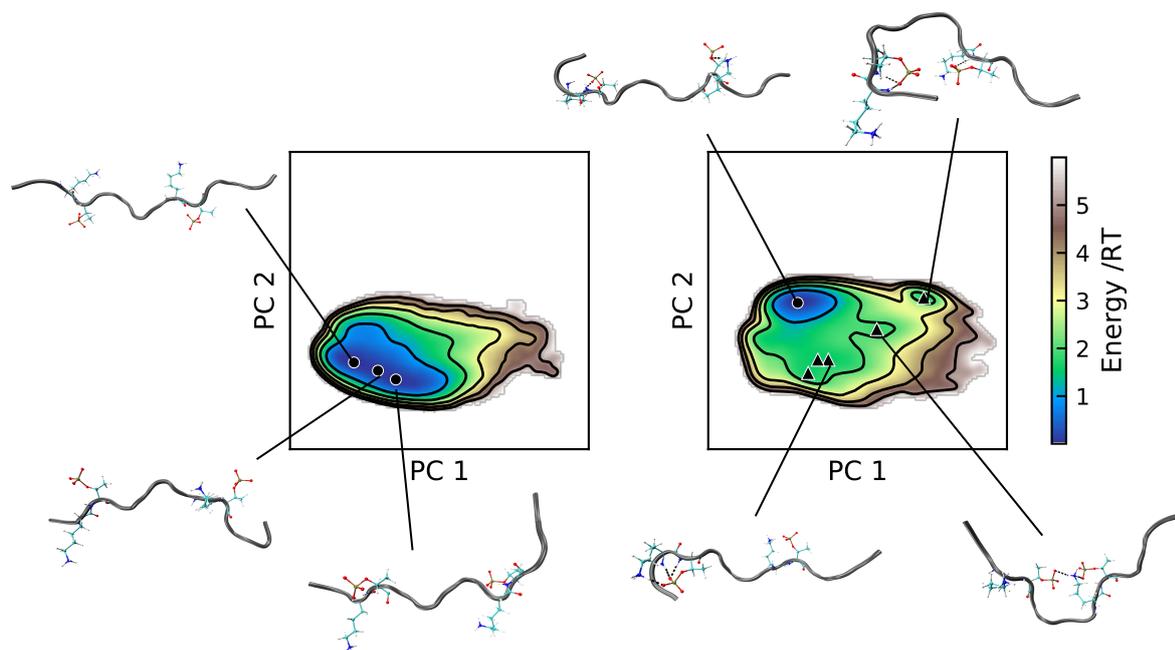
Advancing to the secondary structure, Figure 4 shows that the peptides were mainly irregular, although Tau1 contained much of the polyproline type II (PPII) structure as well. In fact, all peptides contained a significant amount of PPII, as well as a significant amount of bends. The content of the helical structure ( $\alpha$ - and  $3_{10}$ -helix) and  $\beta$ -strands was low in all peptides. Tau1 exhibited the largest differences between the force fields, where A99 produced 16 percentage points more of the PPII structure than C36, which instead contained a more irregular structure. For the other peptides, the differences were smaller. Overall, the peptides only had one significant difference in common, which was a higher content of bends in C36 than A99. Inspecting the content along the sequence, it was evident that it was mostly the same parts of the peptide that were enriched in a certain type of structure in both force fields (see Supplementary Figure S3). However, in C36, the helical content was completely missing from the first ten residues of Stath, which is concerning since the N-terminal region has been shown to possess helical propensity in water, although being mainly disordered [6,30]. Another striking difference between the force fields for Stath is that some residues centered on residues Y21 and Y41 occasionally formed a  $\beta$ -sheet or  $\beta$ -bridge in C36, but not in A99. Notice also that for Tau2, the bend propensity at residues V228–V229 was much higher in C36 than in A99. Since these residues were located right between K225 and pT231, which in C36 formed a stable salt bridge, this suggested that the bend was formed as a result of the salt bridge. Furthermore, for Tau2, NMR data have suggested approximately 40%  $\alpha$ -helical propensity in region A15–R18 [12]. Both A99 and C36 sampled the helical structure in this region, however, to a lower extent than what the experimental data suggested.



**Figure 4.** Average content of different types of the secondary structure in (a) Tau1, (b) Tau2, (c) bCPP, and (d) Stath simulated with Amber ff99SB-ILDN (A99) and CHARMM36m (C36). The legend applies to all panels. The helix includes the  $\alpha$ - $3_{10}$ - and a negligible content of the  $\pi$ -helix, while the  $\beta$ -strand also includes  $\beta$ -bridge. Error bars correspond to errors calculated by block averaging.

### 2.3. Energy Landscapes

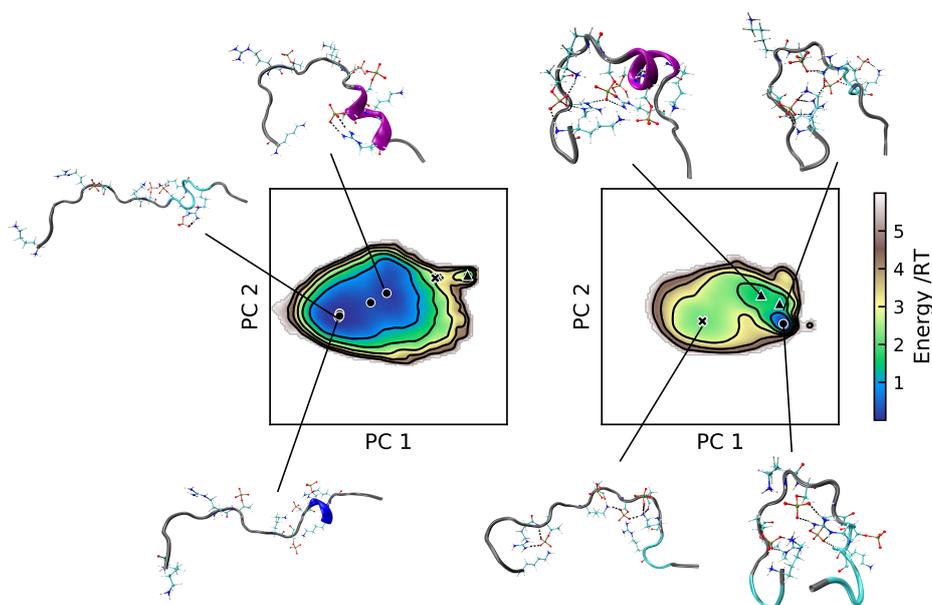
The differences between the force fields in this study is well summarized by the energy landscapes in Figures 5–8. Tau2, bCPP, and Stath all showed a narrower energy landscape in C36, in line with a more restricted conformational ensemble. Tau1, which is rather short and stiff, actually gained a larger conformational landscape in C36, due to sampling more bent conformations in addition to being more stretched out as in A99; see Figure 5. Notice also that in C36, the global minimum, which was the most populated, contained conformations that were not entirely stretched out. Instead, the N-terminal end was folded over, such that a salt bridge was formed between pT175 and the positively charged N-terminus.



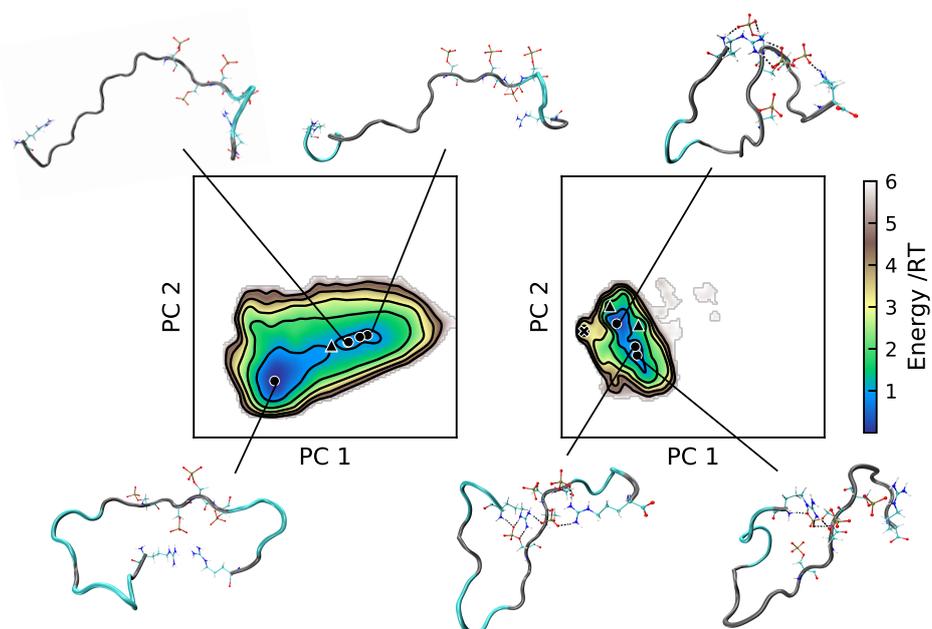
**Figure 5.** Energy landscapes and conformations in selected minima of Tau1. **(Left)** A99; **(right)** C36. The energy landscapes are constructed using the first two components from principal component analysis, using the same basis set for both force fields, such that they are directly comparable. Contour lines are drawn for integer energy levels in the interval  $1 \leq RT \leq 5$ , and the minimum of each basin is represented by a marker: ●: energy  $\leq 1RT$ , ▲:  $\leq 2RT$ . In the conformations, the phosphorylated and positively charged residues are shown explicitly. Dashed black lines represent hydrogen bonds. The peptide conformations are color-coded according to the secondary structure determination in VMD, where silver is irregular (coil) and cyan is turns. The N-terminus of each conformation is the leftmost end.

Although the energy landscapes of Tau2 in A99 and C36 were located in almost the same area, the energy levels differed; see Figure 6. The most populated basin in the C36 simulation was a deep and narrow minimum, while the A99 simulation had a larger area of energy  $\leq 1RT$ , containing several basins, more typical of IDPs. The salt bridges creating more compact conformations were evident in the C36 conformations, while the A99 conformations were more stretched out with fewer salt bridges. Notice that the phosphorylated residues in C36 had a tendency to interact with several positively charged residues simultaneously. In both force fields, a basin minimum with a helical region starting with pS237 and pS238 was found, in line with the secondary structure analysis.

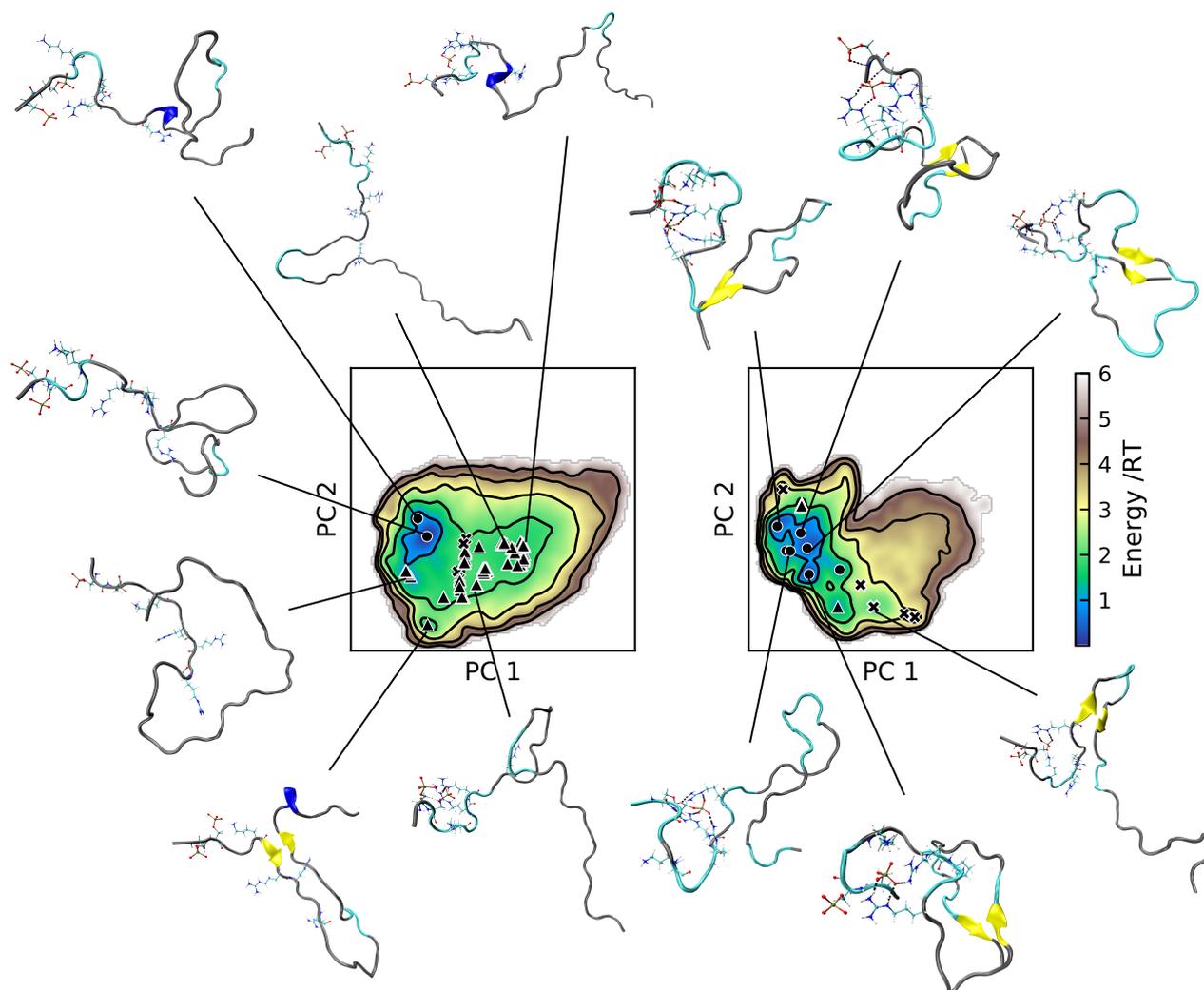
For bCPP, there was indeed many more elongated conformations in the A99 simulation (see Figure 7), and it is clear that what caused the more compact conformations in C36 was the salt bridges between the phosphorylated serines and the arginines. In C36, all depicted conformations contained at least one salt bridge between phosphoserine and arginine, while this was much rarer in A99, explaining why the energy landscapes looked so different. Regarding Stath, comparing the conformations in Figure 8, there were two striking differences. First, there was a higher presence of salt bridges between phosphoserine and positively charged residues in C36, keeping the N-terminal end in a more bent conformation. Secondly, in C36, the  $\beta$ -strand and  $\beta$ -bridge formation between the middle region and C-terminal region detected in Supplementary Figure S3 contributed to making the conformations more compact compared to A99.



**Figure 6.** Energy landscapes and conformations in selected minima of Tau2. (Left) A99; (right) C36. The energy landscapes are constructed using the first two components from principal component analysis, using the same basis set for both force fields, such that they are directly comparable. Contour lines are drawn for integer energy levels in the interval  $1 \leq RT \leq 5$ , and the minimum of each basin is represented by a marker: ●: energy  $\leq 1RT$ , ▲:  $\leq 2RT$ , ✕:  $\leq 3RT$ . In the conformations, the phosphorylated and positively charged residues are shown explicitly. Dashed black lines represent hydrogen bonds. The peptide conformations are color-coded according to the secondary structure determination in VMD, where silver is irregular (coil), cyan is turns, magenta is the  $\alpha$ -helix, and blue is the  $3_{10}$ -helix. The N-terminus of each conformation is the leftmost end.



**Figure 7.** Energy landscapes and conformations in selected minima of bCPP. (Left) A99; (right) C36. The energy landscapes are constructed using the first two components from principal component analysis, using the same basis set for both force fields, such that they are directly comparable. Contour lines are drawn for integer energy levels in the interval  $1 \leq RT \leq 5$ , and the minimum of each basin is represented by a marker: ●: energy  $\leq 1RT$ , ▲:  $\leq 2RT$ , ✕:  $\leq 3RT$ . In the conformations, the phosphorylated and positively charged residues are shown explicitly. Dashed black lines represent hydrogen bonds. The peptide conformations are color-coded according to the secondary structure determination in VMD, where silver is irregular (coil) and cyan is turns. The N-terminus of each conformation is the leftmost end.

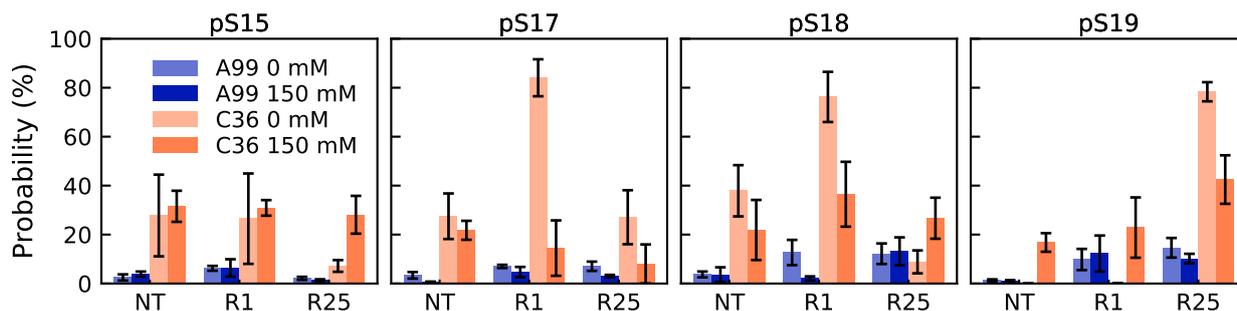


**Figure 8.** Energy landscapes and conformations in selected minima of Stath. (**Left**) A99; (**right**) C36. The energy landscapes are constructed using the first two components from principal component analysis, using the same basis set for both force fields, such that they are directly comparable. Contour lines are drawn for integer energy levels in the interval  $1 \leq RT \leq 5$ , and the minimum of each basin is represented by a marker: ●: energy  $\leq 1RT$ , ▲:  $\leq 2RT$ , ✖:  $\leq 3RT$ . In the conformations, the phosphorylated and positively charged residues are shown explicitly. Dashed black lines represent hydrogen bonds. The peptide conformations are color-coded according to the secondary structure determination in VMD, where silver is irregular (coil), cyan is turns, blue is the  $3_{10}$ -helix, yellow is the  $\beta$ -sheet, and tan is the  $\beta$ -bridge. The N-terminus of each conformation is the leftmost/topmost end.

#### 2.4. Effect of Salt Concentration

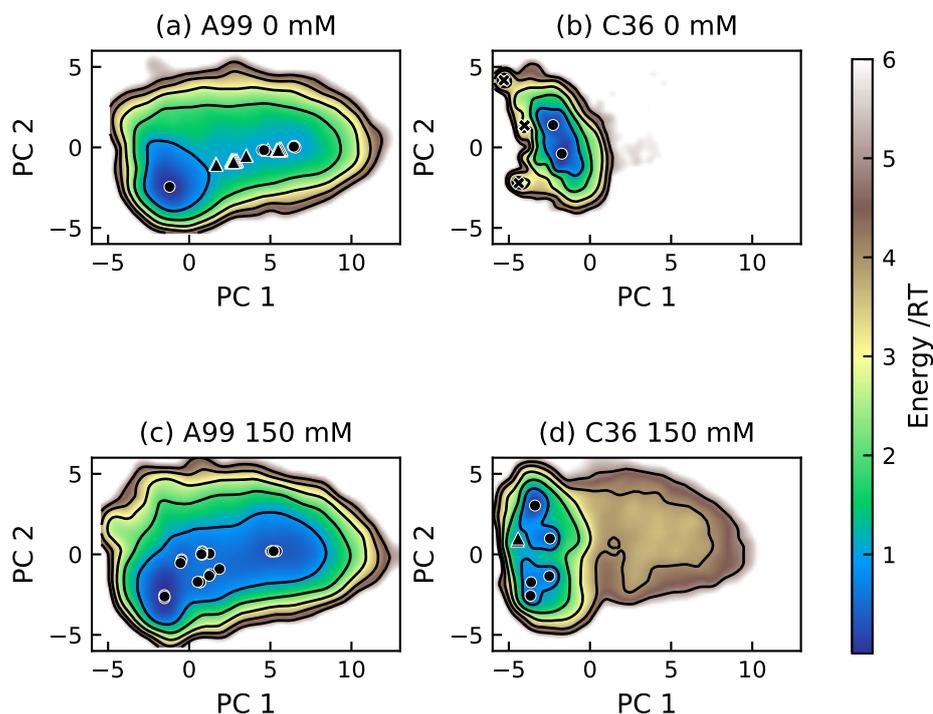
Since the salt bridges formed between phosphorylated and positively charged residues were shown to influence the conformational ensemble, it is of importance to also consider the effect of the screening of the electrostatic interactions. Here, we focused on bCPP, which due to showing the largest differences between force fields and having the highest fraction of charged residues in combination with the largest charge separation (see Supplementary Table S1), was expected to show the largest response to ionic strength. Figure 9 shows that in C36, four of the salt bridges were dramatically reduced upon the addition of 150 mM NaCl; however, the probability of two other salt bridges increased, whereas in A99, only one salt bridge was significantly reduced. At 150 mM salt, the salt-bridging probability was more comparable between A99 and C36, although overall still higher in C36. Supplementary

Figure S3 shows the changes in the contact map upon the addition of 150 mM NaCl for bCPP simulated in A99 and C36. For A99, we clearly saw that the preference for the N-terminal end to be in contact with the phosphorylated and negatively charged region (residues 14–21) diminished. In C36, the strongly conserved R1–pS17 and R1–pS18 contacts were greatly decreased, while the contact of R1 with surrounding residues in the negatively charged region was increased. Hence, this suggested an increased mobility, while still maintaining contact with the negatively charged region. In C36, the cross-diagonal lines also signaled a decrease of the  $\beta$ -sheet; however, the content was relatively low from the beginning.



**Figure 9.** Probability of possible salt bridge interactions for the phosphorylated residues with the N-terminus (NT) and positively charged residues in bCPP, simulated with the two different force fields in the presence of 0 mM or 150 mM NaCl. Error bars corresponds to errors calculated by block averaging.

By comparing the energy landscapes in Figure 10, it is clear that screening of the electrostatic interactions indeed broadened the conformational ensemble, but mainly in C36, which also showed the largest change in salt bridge probability. In C36, the addition of 150 mM NaCl led to the exploration of more stretched out conformations; however, more compact conformations still clearly dominated. A99 also showed an increased probability of visiting more stretched out conformations after the addition of 150 mM NaCl. This shift in the conformational ensemble was also observed in the distributions of  $R_g$  and  $R_{ee}$  shown in Supplementary Figure S4. However, the changes were actually rather small, such that the average values were indistinguishable. Upon the addition of salt, the  $R_g$  changed from  $1.43 \pm 0.03$  nm to  $1.45 \pm 0.03$  nm for A99 and from  $1.08 \pm 0.02$  nm to  $1.08 \pm 0.03$  nm for C36. The changes in  $R_{ee}$  were from  $3.09 \pm 0.15$  nm to  $3.37 \pm 0.13$  nm and from  $1.65 \pm 0.10$  nm to  $1.67 \pm 0.10$  nm, respectively. The effect of salt on the calculated scattering curves was also so small that it could be deemed negligible; see Supplementary Figure S5.



**Figure 10.** Energy landscapes of bCPP simulated with the two force fields Amber ff99SB-ILDN (A99) and CHARMM36m (C36) in the presence of 0 mM or 150 mM NaCl.

### 3. Conclusions

C36 produced more compact conformations of all four peptides, which indeed was expected to be caused mainly by salt bridge stability. In Tau1, the salt bridges pT175–K174 and pT181–K180 were formed without much effect on the overall conformation; however, an additional salt bridge between the N-terminus and pT175 decreased  $R_{ee}$  and  $R_g$  in C36. In Stath, the salt bridges contributed to the discrepancy by restricting the conformation of the first 15 residues, in the same way as previously shown for that fragment studied alone [20]. However, also the  $\beta$ -bridge and  $\beta$ -strand formation between the middle and C-terminal region were shown to contribute to more compact conformations. While C36 produced good results of nonphosphorylated short IDPs, it has been shown to underestimate the size of larger IDPs (>60 residues) [31,32]. Since Stath was 43 residues long, and thus the longest peptide included in this study, it is reasonable to believe that other effects also play a role. That bCPP showed the largest difference between the force fields and Tau1 the smallest implies that the separation between the phosphorylated and positively charged residues controls how much the conformational ensemble is influenced by stable salt bridges. This is in accordance with the importance of considering the level of charge separation for predicting the conformational ensemble of IDPs with a high fraction of charges [33].

When comparing to experimental data, it is important to consider the effect of salt, since most experiments are performed in the presence of buffer and additional salt. In bCPP, the addition of 150 mM NaCl was shown to dramatically reduce the probability of some of the salt bridges in C36, whereas the probability of other salt bridges actually increased. In A99, only one salt bridge was significantly reduced, which suggests that salt bridges still are of importance at 150 mM NaCl. Considering the changes in salt bridge probability for bCPP with salt concentration, it is plausible that the discrepancies between the simulations and experimental reference for Tau2 were caused by nonmatching ionic strength, since the experiments were performed with 50 mM phosphate buffer. At the same time, it can be hard to discern the salt bridges involving close-by residues experimentally, such as for pS237, pS238, K240, and R242.

Despite significant differences in the salt-bridging probability in C36, the effect of salt concentration on the global conformational level, such as  $R_g$  and  $R_{ee}$ , was small enough to be negligible for both force fields. In fact, the calculated form factor was indistinguishable, implying that comparing simulations performed without salt with experimental SAXS data collected at 150 mM NaCl indeed can be valid. Since bCPP is the peptide for which we expected the largest effects of salt concentration, this further strengthens the comparison with SAXS data for Stath collected at 150 mM NaCl, which showed that A99 was in good agreement, while C36 overestimated the level of compaction. Although the effects of ionic strength seem negligible in this study, this is generally not the case. For example, Jin and Gräter needed 350 mM of salt in simulations with A99 to reach experimental agreement for IDPs that are approximately 80 residues long [21], which suggested that also A99 overestimate the strength of salt bridges. Here, both Tau1 and Stath were compared to experimental size estimates, and only C36 was with certainty shown to underestimate the size. Hence, a possible overestimation of salt bridge stability in A99 is not expected to be a major issue for describing the conformational ensemble of the short IDPs studied in this work. This emphasizes the importance of benchmarking against IDPs of different length and sequence when developing and evaluating force fields. While a reduction of the strength of salt bridges appears to be a crucial step in improving the performance of C36, it appears less critical in A99. However, note that this statement is based only on the global conformational properties and that it might be different for studies of dynamics. Based on observations that many force fields have a tendency to overstabilize salt bridges, which seems to be related to side-chain partial charges [22,34–36], we suggest that readjusting the side-chains' partial charges, especially of the phosphorylated residues, is a way of improving the force fields.

Another area which has not been touched upon in this work is the influence of charge regulation and pH. The simulations have been performed with fixed charges in a state corresponding to physiological pH, where the phosphorylated residues have a charge of  $-2e$ . Since the pKa of the phosphorylated residues is around six [37], in reality it can fluctuate between  $-1e$  and  $-2e$ . Recent studies have suggested the importance of the protonation state of phosphorylated residues for molecular interactions [38], hence influencing salt bridge formation and the conformational ensemble. Therefore, this is suggested to be included in future investigations.

Considering the secondary structure, the only general difference between the force fields was a higher content of bends in C36. In Tau2, it was focused on regions between salt-bridging-forming partners, suggesting that highly stable salt bridges can enforce bends depending on the separation between the salt-bridging residues. For Tau2, it was suggested that both force fields underestimated the helical propensity, and in Stath, a lack of helix propensity in the N-terminal regions was concerning for C36. However, to properly assess the performance of force fields regarding the secondary structure, detailed experimental references are important. Hence, we see that NMR experiments of phosphorylated IDPs recording coupling constants, NOEs, and chemical shifts, which capture the effects of both the secondary structure and salt bridges, are an essential part of improving force fields. Since atomistic simulations can be used to carefully detect the secondary structure and salt bridges and their dynamics, it is an important tool in understanding the mechanism behind the regulation of IDP function by phosphorylation, provided that sufficient accuracy of the force fields is achieved.

#### 4. Materials and Methods

Fraction of charged residues and  $\chi$ , a parameter describing how segregated the charged residues are in the sequence [33] were calculated in CIDER [39], by equalizing the phosphorylated residues to other negatively charged residues. The value of  $\chi$  is normalized against the most segregated sequence for that sequence composition, therefore adopting a value in the range 0–1, where 1 corresponds to the most segregated sequence possible.

The simulations listed in Supplementary Table S3 were performed in GROMACS 2018.4 [40–44], using two different force fields: Amber ff99SB-ILDN [23] with the TIP4P-D [24] water model and parameters for the phosphorylated residues from Homeyer et al. [25] and Steinbrecher et al. [26], and CHARMM36m [27] with the CHARMM-modified TIP3P water model [28]. Initial configurations of the peptides were constructed from the sequence as linear chains using Avogadro 1.2.0 [45], optimizing the structure with the auto-optimization tool. Each peptide was solvated in a rhombic dodecahedron box, having a minimum distance between the peptide and the box edges of 1 nm. Sodium ions were added to neutralize the system, and two systems were also simulated with sodium and chloride ions in a concentration corresponding to 150 mM. Periodic boundary conditions were employed in all directions. The equations of motion were integrated using the Verlet leapfrog algorithm [46] with a time step of 2 fs. Nonbonded interactions were treated with a Verlet list cutoff scheme. The short-range interactions were calculated using neighbor lists with cutoff 1 nm or 1.2 nm, for the Amber and CHARMM force fields, respectively. For the CHARMM force field, the Lennard–Jones interactions were switched off smoothly (force-switch) between 1 nm and 1.2 nm. Long-range dispersion corrections were applied to energy and pressure in the case of the Amber force field. Long-range electrostatic interactions were treated by particle mesh Ewald [47] with a cubic interpolation and a 1.6 Å grid spacing. The LINCS algorithm [48] was used to constrain all bond lengths in the case of Amber and only bonds with hydrogen atoms in the case of CHARMM. The solute and solvent were separately coupled to temperature baths at 298 K using the velocity rescaling thermostat [49] with a 0.1 ps relaxation time. Parrinello–Raman pressure coupling [50] was used to keep the pressure at 1 bar, using a 2 ps relaxation time and  $4.5 \times 10^{-5} \text{ bar}^{-1}$  isothermal compressibility.

Energy minimization was performed by the steepest descent algorithm until the system converged within the available machine precision. Initiation of five replicates per system with different starting seeds was performed separately in two steps using position restraints on the peptide. The first step was 500 ps of NVT simulation (constant number of particles, volume, and temperature) performed to stabilize the temperature, followed by the second step of 1000 ps of NPT simulation (constant number of particles, pressure, and temperature) to stabilize the pressure. Production runs of the five replicates per system were performed in the NPT ensemble, for at least 1  $\mu\text{s}$  per replicate. The total simulation time per system is stated in Supplementary Table S3. Energies and coordinates were saved every 10 ps. Supplementary Tables S4 and S5 compile a few differences applied to the salt simulations to reduce the computational time.

### Analysis

The convergence and sampling quality were assessed in the following ways. The time evolution of the  $R_g$  and the  $R_{ee}$  in the simulations were observed for signs of equilibration in the initial stage. Based on this, the first 30 ns were removed from each replicate of bCPP in CHARMM36m and the first 50 ns of each replicate of Tau2 in CHARMM36m before final analysis (see Supplementary Figures S21 and S24). In other systems the equilibration was deemed fast enough to be negligible. The distributions of the  $R_g$  and the  $R_{ee}$  as well as the energy landscapes were compared between replicates, since similarity indicates sufficient sampling. The autocorrelation function and block average error estimates of the  $R_g$  and the  $R_{ee}$  in the concatenated simulation were calculated and observed for an estimate of the correlation time and convergence of the error estimates. All this data is presented in the Supplementary Figures S6–S33. Although some systems showed greater dissimilarity between replicates than desired, based on the assessment of the concatenated trajectory, it was deemed sufficiently sampled to allow for a comparison between the force fields.

$R_g$  and  $R_{ee}$  were calculated using GROMACS 2018.4 [40–44]. Reported error estimates were calculated using block averaging analysis as implemented in the *gmx analyze* routine in GROMACS. Scattering curves were calculated using CRY SOL Version 2.8.3 [51] with the contrast of the hydration shell being  $0.0075 \text{ e}/\text{\AA}^3$  for Amber ff99SB-ILDN+TIP4P-D and

0.02 e/Å<sup>3</sup> for CHARMM36m, as suggested by [29]. The presented curve is the average over 10,000 equally spaced frames. In Supplementary Figure S1 and Table S2, the effect of different contrasts of the hydration shell is shown for Stath. The quality of fit to the experimental curve is computed as:

$$\chi^2(f, c) = N_q^{-1} \sum_{i=1}^{N_q} \left[ \frac{I_{\text{ref}}(q_i) - (f I_{\text{obs}}(q_i) + c)}{\sigma_{\text{ref}}(q_i)} \right]^2, \quad (1)$$

where  $N_q^{-1}$  is the number of points in the reference curve,  $I_{\text{ref}}$  and  $I_{\text{obs}}$  are the reference and observed intensities, respectively, and  $\sigma_{\text{ref}}(q_i)$  is the error associated with each data point of the reference curve. The function was minimized using the Nelder–Mead method [52], as implemented in Scipy [53], using linear interpolation to produce  $I_{\text{obs}}$  at the same  $q$  points as the reference [29]. AUTORG in the ATSAS program [54] was used to determine the  $R_g$  from Guinier analysis. The secondary structure was determined using the DSSP program Version 2.2.1 [55] with an extension to detect the polyproline type II structure [56,57]. The MDTraj Python library Version 1.9.3 [58] was used to calculate contact probability and analyze salt bridges. Contact between two residues was defined as when the shortest distance between two atoms < 0.4 nm. Since salt bridges are formed as a result of hydrogen bonding and electrostatic interactions, they were assessed by analyzing the presence of hydrogen bonds based on the criterion in [59], as implemented in MDTraj. Energy landscapes were calculated following the Campos and Baptista approach [60], with the differences described by Henriques et al. [61]. In short, principal component analysis was applied to the Cartesian coordinates of the backbone atoms of the protein, obtained after translational and rotational least squares fitting on the central structure of the simulation. The conditional free energy was calculated from the probability density function in the representation space constructed by the first two principal components, obtained by Gaussian kernel density estimation. The basins and minima were assigned as described by Campos and Baptista [60]. It is worth noting that the first two components were shown to account for 46–60% of the variance, hence not providing a complete picture of the conformational classes, but at least an overview sufficient for comparison between the force fields. Snapshots from the simulations were produced using VMD 1.9.3 [62–64].

**Supplementary Materials:** The following are available online at [www.mdpi.com/article/10.3390/ijms221810174/s1](http://www.mdpi.com/article/10.3390/ijms221810174/s1).

**Author Contributions:** Conceptualization, M.S. and E.R.; formal analysis, E.R.; funding acquisition, M.S.; investigation, E.R.; methodology, M.S. and E.R.; project administration, M.S.; resources, M.S.; supervision, M.S.; validation, M.S. and E.R.; visualization, E.R.; writing, M.S. and E.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** Financial support was given by the Royal Physiographic Society in Lund and the Bertil Andersson foundation.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is included in the article or Supplementary Materials.

**Acknowledgments:** The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the Center for Scientific and Technical Computing at Lund University (LUNARC).

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

A99	Amber ff99SB-ILDN with TIP4P-D water
C36	CHARMM36m with CHARMM-modified TIP3P water
FRET	Fluorescence resonance energy transfer
IDP	Intrinsically disordered protein
NMR	Nuclear magnetic resonance
PPII	polyproline type II
R <sub>g</sub>	Radius of gyration
R <sub>ee</sub>	End-to-end distance
SAXS	Small-angle X-ray scattering

### References

- Dunker, A.; Lawson, J.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59. [[CrossRef](#)]
- Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [[CrossRef](#)]
- Fisher, C.K.; Stultz, C.M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426–431. [[CrossRef](#)] [[PubMed](#)]
- Johnson, L.N.; Lewis, R.J. Structural Basis for Control by Phosphorylation. *Chem. Rev.* **2001**, *101*, 2209–2242. [[CrossRef](#)] [[PubMed](#)]
- Gong, C.X.; Iqbal, K. Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease. *Curr. Med. Chem.* **2008**, *15*, 2321–2328. [[CrossRef](#)]
- Raj, P.A.; Johnsson, M.; Levine, M.J.; Nancollas, G.H. Salivary statherin. Dependence on sequence, charge, hydrogen bonding potency, and helical conformation for adsorption to hydroxyapatite and inhibition of mineralization. *J. Biol. Chem.* **1992**, *267*, 5968–5976. [[CrossRef](#)]
- Makrodimitris, K.; Masica, D.L.; Kim, E.T.; Gray, J.J. Structure Prediction of Protein–Solid Surface Interactions Reveals a Molecular Recognition Motif of Statherin for Hydroxyapatite. *J. Am. Chem. Soc.* **2007**, *129*, 13713–13722. [[CrossRef](#)] [[PubMed](#)]
- De Kruif, C.G.; Holt, C. Casein Micelle Structure, Functions and Interactions. In *Advanced Dairy Chemistry—1 Proteins: Part A/Part B*; Fox, P.F., McSweeney, P.L.H., Eds.; Springer: Boston, MA, USA, 2003; pp. 233–276. [[CrossRef](#)]
- Martin, E.W.; Holehouse, A.S.; Grace, C.R.; Hughes, A.; Pappu, R.V.; Mittag, T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J. Am. Chem. Soc.* **2016**, *138*, 15323–15335. [[CrossRef](#)]
- Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J.D. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **2010**, *18*, 494–506. [[CrossRef](#)]
- Chin, A.; Toptygin, D.; Elam, W.; Schrank, T.; Hilser, V. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. *Biophys. J.* **2016**, *110*, 362–371. [[CrossRef](#)]
- Schwalbe, M.; Kadavath, H.; Biernat, J.; Ozenne, V.; Blackledge, M.; Mandelkow, E.; Zweckstetter, M. Structural Impact of Tau Phosphorylation at Threonine 231. *Structure* **2015**, *23*, 1448–1458. [[CrossRef](#)]
- RFarrell, H.; Qi, P.; Wickham, E.; Unruh, J. Secondary Structural Studies of Bovine Caseins: Structure and Temperature Dependence of  $\beta$ -Casein Phosphopeptide (1–25) as Analyzed by Circular Dichroism, FTIR Spectroscopy, and Analytical Ultracentrifugation. *J. Protein Chem.* **2002**, *21*, 307–321. [[CrossRef](#)]
- Brister, M.A.; Pandey, A.K.; Bielska, A.A.; Zondlo, N.J. OGlcnAcylation and Phosphorylation Have Opposing Structural Effects in tau: Phosphothreonine Induces Particular Conformational Order. *J. Am. Chem. Soc.* **2014**, *136*, 3803–3816. [[CrossRef](#)] [[PubMed](#)]
- Cragnell, C.; Rieloff, E.; Skepö, M. Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions. *J. Mol. Biol.* **2018**, *430*, 2478–2492. [[CrossRef](#)]
- Sieradzan, A.K.; Bogunia, M.; Mech, P.; Ganzynkiewicz, R.; Giełdoń, A.; Liwo, A.; Makowski, M. Introduction of Phosphorylated Residues into the UNRES Coarse-Grained Model: Toward Modeling of Signaling Processes. *J. Phys. Chem. B* **2019**, *123*, 5721–5729. [[CrossRef](#)] [[PubMed](#)]
- Sieradzan, A.K.; Korneev, A.; Begun, A.; Kachlishvili, K.; Scheraga, H.A.; Molochkov, A.; Senet, P.; Niemi, A.J.; Maisuradze, G.G. Investigation of Phosphorylation-Induced Folding of an Intrinsically Disordered Protein by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 3203–3220. [[CrossRef](#)] [[PubMed](#)]
- Chong, S.H.; Chatterjee, P.; Ham, S. Computer Simulations of Intrinsically Disordered Proteins. *Annu. Rev. Phys. Chem.* **2017**, *68*, 117–134. [[CrossRef](#)] [[PubMed](#)]
- Huang, J.; MacKerell, A.D. Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2018**, *48*, 40–48. [[CrossRef](#)]
- Rieloff, E.; Skepö, M. Phosphorylation of a Disordered Peptide—Structural Effects and Force Field Inconsistencies. *J. Chem. Theory Comput.* **2020**, *16*, 1924–1935. [[CrossRef](#)]
- Jin, F.; Gräter, F. How multisite phosphorylation impacts the conformations of intrinsically disordered proteins. *PLoS Comput. Biol.* **2021**, *17*, e1008939. [[CrossRef](#)]

22. Ahmed, M.C.; Papaleo, E.; Lindorff-Larsen, K. How well do force fields capture the strength of salt bridges in proteins? *PeerJ* **2018**, *6*, e4967. [[CrossRef](#)]
23. 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* **2010**, *78*, 1950–1958. [[CrossRef](#)] [[PubMed](#)]
24. 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. [[CrossRef](#)]
25. Homeyer, N.; Horn, A.H.C.; Lanig, H.; Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.* **2006**, *12*, 281–289. [[CrossRef](#)] [[PubMed](#)]
26. Steinbrecher, T.; Latzer, J.; Case, D.A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405–4412. [[CrossRef](#)]
27. 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* **2016**, *14*, 71–73.
28. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [[CrossRef](#)]
29. Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* **2018**, *430*, 2521–2539. [[CrossRef](#)] [[PubMed](#)]
30. Naganagowda, G.A.; Gururaja, T.L.; Levine, M.J. Delineation of Conformational Preferences in Human Salivary Statherin by <sup>1</sup>H, <sup>31</sup>P NMR and CD Studies: Sequential Assignment and Structure-Function Correlations. *J. Biomol. Struct. Dyn.* **1998**, *16*, 91–107. [[CrossRef](#)]
31. Robustelli, P.; Piana, S.; Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4758–E4766. [[CrossRef](#)] [[PubMed](#)]
32. Chan-Yao-Chong, M.; Deville, C.; Pinet, L.; van Heijenoort, C.; Durand, D.; Ha-Duong, T. Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data. *Biophys. J.* **2019**, *116*, 1216–1227. [[CrossRef](#)]
33. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13392–13397. [[CrossRef](#)] [[PubMed](#)]
34. Piana, S.; Lindorff-Larsen, K.; Shaw, D.E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–L49. [[CrossRef](#)]
35. Debiec, K.T.; Gronenborn, A.M.; Chong, L.T. Evaluating the Strength of Salt Bridges: A Comparison of Current Biomolecular Force Fields. *J. Phys. Chem. B* **2014**, *118*, 6561–6569. [[CrossRef](#)]
36. Best, R.B. Atomistic Force Fields for Proteins. In *Biomolecular Simulations: Methods and Protocols*; Bonomi, M., Camilloni, C., Eds.; Springer: New York, NY, USA, 2019; pp. 3–19. [[CrossRef](#)]
37. Bienkiewicz, E.A.; Lumb, K.J. Random-coil chemical shifts of phosphorylated amino acids. *J. Biomol. NMR* **1999**, *15*, 203–206. [[CrossRef](#)]
38. Kawade, R.; Kuroda, D.; Tsumoto, K. How the protonation state of a phosphorylated amino acid governs molecular recognition: insights from classical molecular dynamics simulations. *FEBS Lett.* **1999**, *594*, 903–912. [[CrossRef](#)] [[PubMed](#)]
39. Holehouse, A.S.; Das, R.K.; Ahad, J.N.; Richardson, M.O.G.; Pappu, R.V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **2017**, *112*, 16–21. [[CrossRef](#)] [[PubMed](#)]
40. Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. [[CrossRef](#)]
41. 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. [[CrossRef](#)]
42. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; van der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854. [[CrossRef](#)]
43. Páll, S.; Abraham, M.J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 3–27.
44. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [[CrossRef](#)]
45. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17. [[CrossRef](#)] [[PubMed](#)]
46. Hockney, R.W.; Eastwood, J.W. *Computer Simulation Using Particles*; McGraw-Hill: New York, NY, USA, 1981.
47. 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*, 10089–10092. [[CrossRef](#)]
48. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [[CrossRef](#)]
49. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [[CrossRef](#)]

50. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [[CrossRef](#)]
51. Svergun, D.; Barberato, C.; Koch, M.H.J. CRY SOL—A Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773. [[CrossRef](#)]
52. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
53. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
54. Franke, D.; Petoukhov, M.V.; Konarev, P.V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.D.T.; Kikhney, A.G.; Hajizadeh, N.R.; Franklin, J.M.; Jeffries, C.M.; et al. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. [[CrossRef](#)]
55. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [[CrossRef](#)]
56. 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*, e18401. [[CrossRef](#)]
57. Chebrek, R.; Leonard, S.; de Brevern, A.G.; Gelly, J.C. PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* **2014**, *2014*, bau102. [[CrossRef](#)] [[PubMed](#)]
58. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [[CrossRef](#)]
59. Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L.Å.; Hirsch, T.K.; Ojamäe, L.; Glatzel, P.; et al. The Structure of the First Coordination Shell in Liquid Water. *Science* **2004**, *304*, 995–999. [[CrossRef](#)] [[PubMed](#)]
60. Campos, S.R.R.; Baptista, A.M. Conformational Analysis in a Multidimensional Energy Landscape: Study of an Arginylglutamate Repeat. *J. Phys. Chem. B* **2009**, *113*, 15989–16001. [[CrossRef](#)]
61. 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. [[CrossRef](#)] [[PubMed](#)]
62. Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
63. Stone, J. An Efficient Library for Parallel Ray Tracing and Animation. Master's Thesis, Computer Science Department, University of Missouri-Rolla, Rolla, MO, USA, 1998.
64. Frishman, D.; Argos, P. Knowledge-based secondary structure assignment. *Proteins* **1995**, *23*, 566–579. [[CrossRef](#)]