

RESEARCH ARTICLE

Effects of forcefield and sampling method in all-atom simulations of inherently disordered proteins: Application to conformational preferences of human amylin

Enxi Peng, Nevena Todorova, Irene Yarovsky*

School of Engineering, RMIT University, Melbourne, Victoria, Australia

* irene.yarovsky@rmit.edu.au



OPEN ACCESS

Citation: Peng E, Todorova N, Yarovsky I (2017) Effects of forcefield and sampling method in all-atom simulations of inherently disordered proteins: Application to conformational preferences of human amylin. PLoS ONE 12(10): e0186219. <https://doi.org/10.1371/journal.pone.0186219>

Editor: L. Michel Espinoza-Fonseca, University of Minnesota Twin Cities, UNITED STATES

Received: January 16, 2017

Accepted: September 27, 2017

Published: October 12, 2017

Copyright: © 2017 Peng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by Australian Research Council (<http://www.arc.gov.au>) Discovery grant DP140101888 and DP170100511 awarded to IY. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Although several computational modelling studies have investigated the conformational behaviour of inherently disordered protein (IDP) amylin, discrepancies in identifying its preferred solution conformations still exist between various forcefields and sampling methods used. Human islet amyloid polypeptide has long been a subject of research, both experimentally and theoretically, as the aggregation of this protein is believed to be the lead cause of type-II diabetes. In this work, we present a systematic forcefield assessment using one of the most advanced non-biased sampling techniques, Replica Exchange with Solute Tempering (REST2), by comparing the secondary structure preferences of monomeric amylin in solution. This study also aims to determine the ability of common forcefields to sample a transition of the protein from a helical membrane bound conformation into the disordered solution state of amylin. Our results demonstrated that the CHARMM22* forcefield showed the best ability to sample multiple conformational states inherent for amylin. It is revealed that REST2 yielded results qualitatively consistent with experiments and in quantitative agreement with other sampling methods, however far more computationally efficiently and without any bias. Therefore, combining an unbiased sampling technique such as REST2 with a vigorous forcefield testing could be suggested as an important step in developing an efficient and robust strategy for simulating IDPs.

Introduction

With advances in computational power, the use of detailed atomistic simulations of biological systems has led to a level of understanding that is not achievable by experimental techniques alone.[1–5] Broadly speaking, there are two important requirements that must be satisfied in order to accurately describe the behaviour of biomolecules using computational modelling.[6] First, it is important to have an appropriate forcefield that accurately represents the interactions between atoms within a biomolecular system. Second, sufficient sampling of the potential energy landscape is required to obtain accurate description of the kinetic and thermodynamic properties of the system. The latter of the two is currently one of the main drawbacks of

classical molecular dynamics (MD) applications to biomolecular systems. The sampling deficiency is omnipresent when simulating intrinsically disordered proteins (IDP), since it is essential for these systems to be able to efficiently cross energy barriers between local energy minima.[7] Currently, comprehensive conformational sampling using brute-force MD alone is only achievable on specially designed hardware like Anton designed and purpose build by D.E. Shaw Research.[8]

One important example of an intrinsically disordered protein is the human islet polypeptide (hIAPP), or amylin, which is co-expressed and co-secreted alongside insulin in the pancreatic β -cells. However, when functional amylin proteins misfold they convert into insoluble amyloidogenic fibril-like structures.[9] This disorder is classified as a form of metabolic conformational disease, which had been strongly implicated in the development of type-II diabetes mellitus. [10, 11] Detailed understanding of the conversion process of functional proteins into toxic amyloid fibrils could aid the development of targeted medicinal applications and novel therapeutics. Nonetheless, the structure of monomeric amylin has been extensively studied using experimental techniques; including circular dichroism and solid-state NMR which have shown amylin can adopt a variety of random coil, β -hairpin and α -helical structures.[12–14] Nanga et al. determined a helix-turn-helix conformation of amylin when bound to a micelle (Fig 1A).[15] However, Goldsbury et al. showed the aggregation of amylin to amyloidogenic fibrils involved a conformational change from random coil (Fig 1B) to β -sheet/ α -helical structures. Therefore, an appropriate computational model, including both the conformational sampling method and forcefield, must be capable to reflect these conformational ranges.[12]

As with any protein, it is necessary to gain a comprehensive understanding of its preferred conformational states in solution and classical MD simulations are typically the tool to be used, however a comprehensive conformational sampling is not achievable using brute-force MD, without sacrificing a great amount of time and computational resources.[16] As a result, a range of enhanced sampling methods have been developed in the past decade to overcome the limitations of the brute-force approach including replica exchange molecular dynamics (REMD [17] or T-REMD [18]), bias-exchange metadynamics (BEMD)[19] and, more recently, replica exchange with solute tempering (REST2).[20] In the REMD method, several non-interacting replicas of the same system are simulated at different temperatures. At selected exchange times, a Monte Carlo exchange is performed between the replicas, with the exchange of configuration accepted or rejected based on a Metropolis acceptance criterion. [17, 21] This enables the higher temperature replicas to cross energy barriers, while the low temperature replicas sample larger conformational space. The main drawback of REMD is that a large number of replicas are required to obtain effective sampling, which can be computationally expensive. Another advanced sampling method, bias-exchange metadynamics (BEMD), relies on history-dependent biases to overcome free-energy barriers. In this approach, the choice of

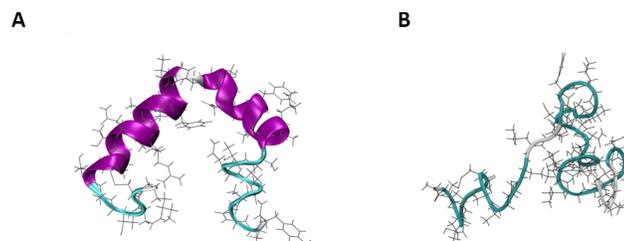


Fig 1. Cartoon/licorice representation of the starting structures of human amylin used in this work. (A) NMR micelle bound structure PDB code 2L86.[15]. (B) Unfolded random coil conformation (taken from our preliminary assessment of amylin).

<https://doi.org/10.1371/journal.pone.0186219.g001>

collective variables is crucial for maintaining the accuracy of system modelled; if slow changing degrees of freedom are not included, inaccurate systematic biases in free-energy can occur. [17, 22] A relatively new advanced sampling technique based on replica exchange with solute tempering is designed to increase the efficiency of protein sampling in aqueous solution by transforming the Hamiltonian of each replica (instead of using different temperatures), and is also known as Hamiltonian-REM. Liu et al. developed this method to only modulate the solute in the simulation, thus increases in energy are only applied to part of the system. This makes replica exchange with solute tempering (revised and renamed as REST2 in [23] capable of advanced sampling using a small number of replicas, and therefore is more computationally efficient relative to other replica exchange methods. [20, 24]

The conformational behaviour of monomeric amylin has been previously investigated using different advanced sampling techniques. Most recently, Hoffman et al. used BEMD to study the forcefield effects on the structure and dynamics of rat and human amylin to determine which forcefield was most suitable to model the species-related conformational differences.[5] Their work highlighted the structural preferences (biases) of specific forcefields, where GROMOS forcefields, in general, exhibited a conformational bias towards hair-pin structures, CHARMM27 towards α -helices, OPLS-AA/L towards random coil, while Amberff03w and CHARMM22* provided a balance between secondary structures consistent with available experimental literature.[5] Furthermore, Zerze et al. provided a comparative analysis of the hIAPP conformational sampling using both T-REMD and BEMD to elucidate the differences between the two methodologies.[4] Results from this study determined that both techniques yielded consistent results based on the free-energy and secondary structure analysis. However, the analyses highlighted that the selection of collective variables in BEMD has to be taken with caution, as lower helical propensity resulted from BEMD compared to T-REMD sampling due to the exclusion of residues 2–7 from the α -RMSD bias. Nonetheless, the studies showed the ability of BEMD and T-REMD to efficiently sample the conformational landscape of intrinsically disordered proteins and provided the necessary benchmarking data for an independent assessment of REST2, which can alleviate the computational expenses and biases associated with T-REMD and BEMD [4]. In this work we used the membrane bound conformation of amylin as starting structure to determine the ability of common forcefields to take the protein out of the helical conformation into the disordered solution state as seen in experiment [12]. Additional simulations starting from the unfolded random coil conformation of amylin were also performed to enhance conformational sampling and elucidate any forcefield related biases. We systematically compare the conformational preferences exhibited by these forcefields using brute-force MD and non-biased sampling techniques such as REST2, with other similar simulations studies [4, 5].

Methodologies

Classical MD simulations

All simulations were carried out using the GROMACS 5.0.5 software package.[25] Classical “brute-force” MD and REST2 simulations were performed on two different starting structures, the native NMR structure of hIAPP (PDB code 2L86) [15] and unfolded random coil using five different forcefields, listed in Table 1. All forcefields used in this study are available in the GROMACS simulations package. The peptide was modelled in zwitterionic form with NH^{3+} and COO^- as termini. The protein was enclosed in a periodic box of 65 Å x 65 Å x 65 Å size, and solvated with 8900 TIP3P, TIP3SP or SPC waters in accord with the forcefield applied to achieve the water density of 1 g/cm³. [3] To investigate the role of the water model in the protein conformational dynamics, the CHARMM22* and CHARMM27 simulations were also

Table 1. Total simulation times collected for each forcefield employed and respective water models. Convergence of the REST2 simulations was determined using cluster analysis, whilst brute-force MD simulation was considered to reach equilibrium when the system energies and backbone RMSD had plateaued. CHARMM22/CMAP is designated as CHARMM27.

Forcefield (FF)	Water Model	Simulation Length (ns)			
		Folded		Unfolded	
		MD	REST2	MD	REST2
AMBER99SB*-ILDN [31]	TIP3P	600	100	500	30
GROMOS96 54a7 [32]	SPC	600	200	500	30
CHARMM36 [33]	TIP3SP	500	100	500	30
CHARMM22* [30]	TIP3P/TIP3SP	500	80	500	30
CHARMM27 [34]	TIP3P/TIP3SP	500	80	500	30

<https://doi.org/10.1371/journal.pone.0186219.t001>

repeated with the modified TIP3P water model (TIP3SP). Two Cl⁻ counter ions were added to the system to neutralise the overall +2 charge. Van der Waals and electrostatic interactions were truncated at 1 nm with long-range electrostatics calculated by Particle Mesh Ewald summation method.[26] Energy minimisation was initially carried out to remove any steric clashes using steepest descent algorithm until convergence was achieved at 250 kJ/mol. This was followed by three equilibration stages: (1) 100 ps NVT (constant volume/temperature) simulation to relax the system, (2) 2 ns simulation having the solute position restrained to relax the solvent around the protein and (3) 10 ns of un-restrained NPT (constant pressure/temperature) simulation. Constant pressure of 1 bar was achieved by coupling the system using the Berendsen barostat, with coupling constant of 2.0 ps and reference pressure of 1.0 bar,[27] and constant temperature of 300 K and coupling constant of 0.1 ps was achieved using the v-rescale [28] method for the course of the simulation. The LINCS algorithm [29] was applied to constrain the bond lengths to their equilibrium values which enabled a simulation time-step of 2 fs to be used. Following this preparation, the systems were subjected to data collection by brute force MD (for comparison purposes) and REST2 simulations, described below and summarised in Table 1. Two concurrent simulations were run with different starting velocities for the brute force MD simulations.

REST2 simulations

For the REST2 simulations, the equilibrated conformation of amylin was used as a starting structure with 16 replicas for each forcefield, except for CHARMM36, where 24 replicas were used (the reason for this is discussed later). Van der Waals and electrostatic interactions were kept consistent with the classical MD settings described above.

The free-energy perturbation and replica exchange code within GROMACS was used to implement the REST2 methodology as outlined by Terakawa et al.[20] and re-examined by the original REST developers, Berne et al. [23], also known as REST2. The effective temperatures for each system were set within the interval of 300 to 600 K. The average probability, λ , for an attempted exchange was approximately 0.5, using a series of 1 ns REST runs, and an exchange was attempted every 1000 MD steps. The total simulation times for each forcefield using REST2 are listed in Table 1. Each REST2 simulation was determined as converged when the clustering of the neutral (unbiased) replica trajectory reached a plateau (i.e. new significant clusters were no longer forming). This approach has been applied in other similar studies using REST2.[35] All analyses of the REST2 simulations were conducted on the neutral replica.

Data analysis. Analyses of the simulation data was performed on the final 50 ns of the converged trajectory for the MD simulations, and 40 ns for the neutral REST2 replica. The

convergence of the REST2 simulations was confirmed by monitoring the development of new highly populated clusters over the last 40 ns of simulations. [35] The convergence of the MD simulations was determined when the system energies and RMSD of the protein backbone had reached a plateau. Here, the plateauing RMSD was used as a criterion for stopping the MD simulation runs considered conformationally trapped and not useful to continue. Due to the nature of classical molecular dynamics simulations, it is not feasible to match MD simulated time to that of REST2, thus classical MD simulation results are only indicative of the limitation of this sampling method. Clustering analysis using the LINKAGE method [36] was first applied to determine the most populated structures during the equilibrated period of the simulation. To obtain the most populated clusters, where ~50% of the sampled structures are found in the top 3 clusters, a different RMSD cut-off was applied for each sampling method. Specifically, a cut-off of 2 Å was used for the REST2 simulations to capture a high variation of conformations sampled, compared to the 1.5 Å cut-off used for the more conformationally restricted brute-force MD generated trajectories. The protein secondary structure content sampled in each simulation was calculated using the STRIDE algorithm in VMD.[37] The average number of residues that adopted a particular secondary structure element (coils, turns, helices and β -strands) was calculated. The three types of helical structures, α -helix, π -helix and 3_{10} -helix were summed together as the helical group. Similarly, strands were defined by the combination of β -bridges and β -sheets. In addition, the software code PLUMED 2.0 [38] was used to construct conformational free-energy landscape sampled by each forcefield starting from the folded conformation using the REST2 and MD (shown in Supporting Information S3 Fig).

Results and discussion

Conformational dynamics

Clustering and secondary structure analyses were performed over the equilibrated period of both brute-force MD and REST2 simulations starting from the folded (mostly helical) and unfolded (random coil) conformations. The results are presented in Figs 2 and 3, respectively. The α -helical versus β -strand content sampled by each forcefield was also determined using PLUMED, (Supporting Information S3 Fig). The MD simulations starting from the folded conformation using the AMBER99SB*-ILDN forcefield exhibited stable conformations, most populated cluster having 85% of equilibrium structures, with some helical content between residues 5–16 and coils along the amyloidogenic C-terminus. Similarly, the simulations starting from the unfolded conformation produced a single cluster (100% population) albeit with a slightly less helical content (~8 residues) with short helical segments along both amyloidogenic sections of the protein. In contrast, the REST2 simulations with the AMBER99SB*-ILDN forcefield, exhibited mostly disordered states dominated by turn and coil conformations, irrespective of the starting structure. Different conformational behaviour was observed in the simulations using the GROMOS96 54a7 forcefield. The brute-force MD simulation starting from the folded amylin conformation resulted in a stable helical structure (~67% population), with ~16 residues on average forming helices and the REST2 forming coils and turns. While the simulation starting from an unfolded amylin showed a 100% dominant cluster composed of β -hairpin-like conformations, as seen in the secondary structure results in Fig 3, where ~24 residues were in the β -strand conformation. Interestingly, this observation was reversed in the REST2 simulations with GROMOS96 54a7 forcefield, albeit the overall reduced β -strand/helical structure content. The two most populated clusters from the simulations starting from the folded structure consisted of a ~6 residue in β -strand conformation, while the remaining clusters with lower populations contained some helical (<5 residues) and random coil conformations. The REST2 simulations starting from the unfolded structure exhibited ~46% of

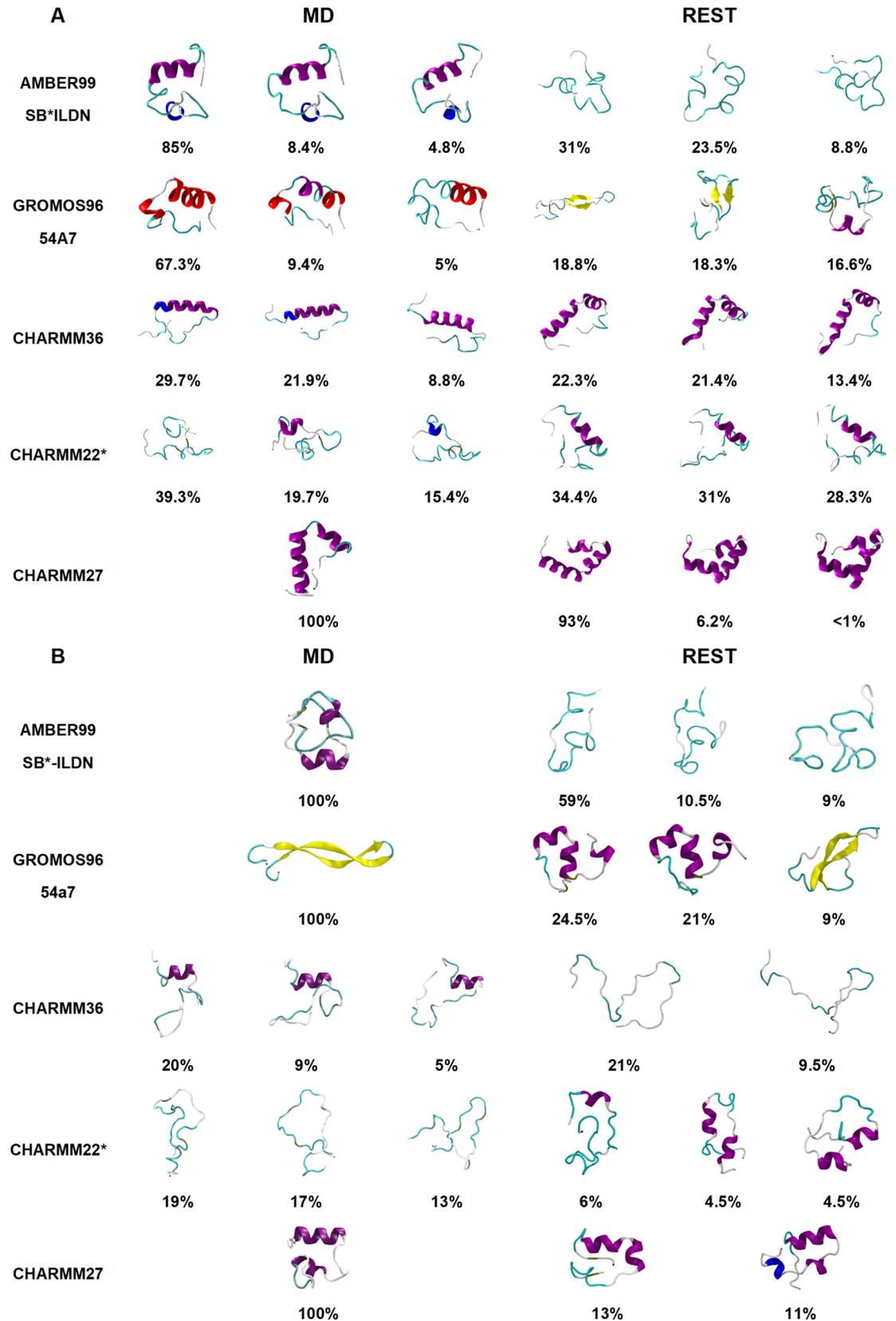


Fig 2. Clustering analysis results from MD and REST2 simulations starting from (A) folded NMR structure and (B) disordered (unfolded) state of amylin. Top clusters for each simulation are selected to represent the most common conformations of the equilibrated trajectory. The protein secondary structure is represented as cartoon with the α -helix coloured in purple, 3-10-helix in blue, turn in cyan and coil in white and extended β -sheet in yellow.

<https://doi.org/10.1371/journal.pone.0186219.g002>

structures with shorter helices spread across ~ 10 residues along the amyloidogenic segments of the protein, whereas there were only low populated clusters exhibiting some β -structuring less than 5 residues on average. The CHARMM forcefield MD simulations starting from the folded conformation presented mostly helical conformations, with the exception of CHARMM22*. The CHARMM36 simulations displayed clusters with extended helices and flexible termini where on average ~ 20 residues were in a helical conformation during the equilibrated period while CHARMM27 produced a single cluster resembling the NMR micelle bound configuration with ~ 13 residues in helical conformation. The brute-force MD trajectories using CHARMM22* forcefield exhibited short 3–5 residue helical segments, present in 60% of the clusters formed, whilst the simulations starting from an unfolded structure remained fully disordered. The CHARMM36 simulations starting from the unfolded structure exhibited a small helical formation ~ 6 residues on average, while the CHARMM27 forcefield was able to fold back the protein into a mostly helical structure with ~ 16 residues adopting helix-like conformations. This result illustrates that brute-force MD is unable to overcome the helical propensities of the CHARMM27 forcefield, as seen in other similar studies [4, 5]. In the REST2 simulations starting from the folded structure using the CHARMM22* forcefield, the protein formed semi-helical conformations similar to that of brute-force MD results, with a slightly longer helical segment along residues 6–16 and random coil formation along residues 20–29; whilst the unfolded REST2 simulations yielded largely random coil conformations with short 3–5 residue helical segments. Similarly, the brute-force MD, the REST2 simulations starting from folded amylin with CHARMM27 and CHARMM36 forcefields had a majority of the structures forming helix-turn-helix conformations. When CHARMM36 was used to simulate 24 replicas rather than the standard 16 (due to poor exchange rates at higher effective temperature replicas), the highly helical secondary structure was still present which may be indicative of the limitations of the conformational sampling. Conversely, REST2 simulations starting from the disordered state of amylin using the same two CHARMM forcefields resulted in less helical structure formation, with CHARMM36 forming entirely random coils, and CHARMM27 forming ~ 11 residue helices as opposed to the folded (helical) simulations. These variances between clusters formed by changing the starting structure illustrate two things. Firstly, in brute-force MD, even by using a fully disordered starting structure, certain forcefields (GROMOS96 54a7 and CHARMM27) showed conformational bias towards a specific secondary structure. Secondly, When REST2 is applied to the unfolded starting structure, previously observed forcefield biases were alleviated across all three CHARMM forcefields tested, which highlights the importance of proper sampling with any forcefield.

Overall, the additional simulations starting from the unfolded structure provided valuable observations into the conformational biases of some forcefields and ways to alleviate them. The brute-force MD results suggested that even though starting from an unfolded conformation, GROMOS96 54a7 favoured the β -hairpin conformation, while CHARMM27 formed the helix-turn-helix structure resembling the micelle bound NMR conformation. AMBER99SB*-ILDN produced a structure with helical content at either terminus, while CHARMM36 produced a short 3–5 residue helical segment and random coils for the REST2 of the residues. Alternatively, CHARMM22* maintained an overall disordered conformation. REST2 was successful at preventing the protein from being trapped in low energy states and was able to sample a wider conformational space. However, when REST2 was applied to an already folded structure in conjunction with GROMOS96 54a7 and CHARMM27 it was unable to take the

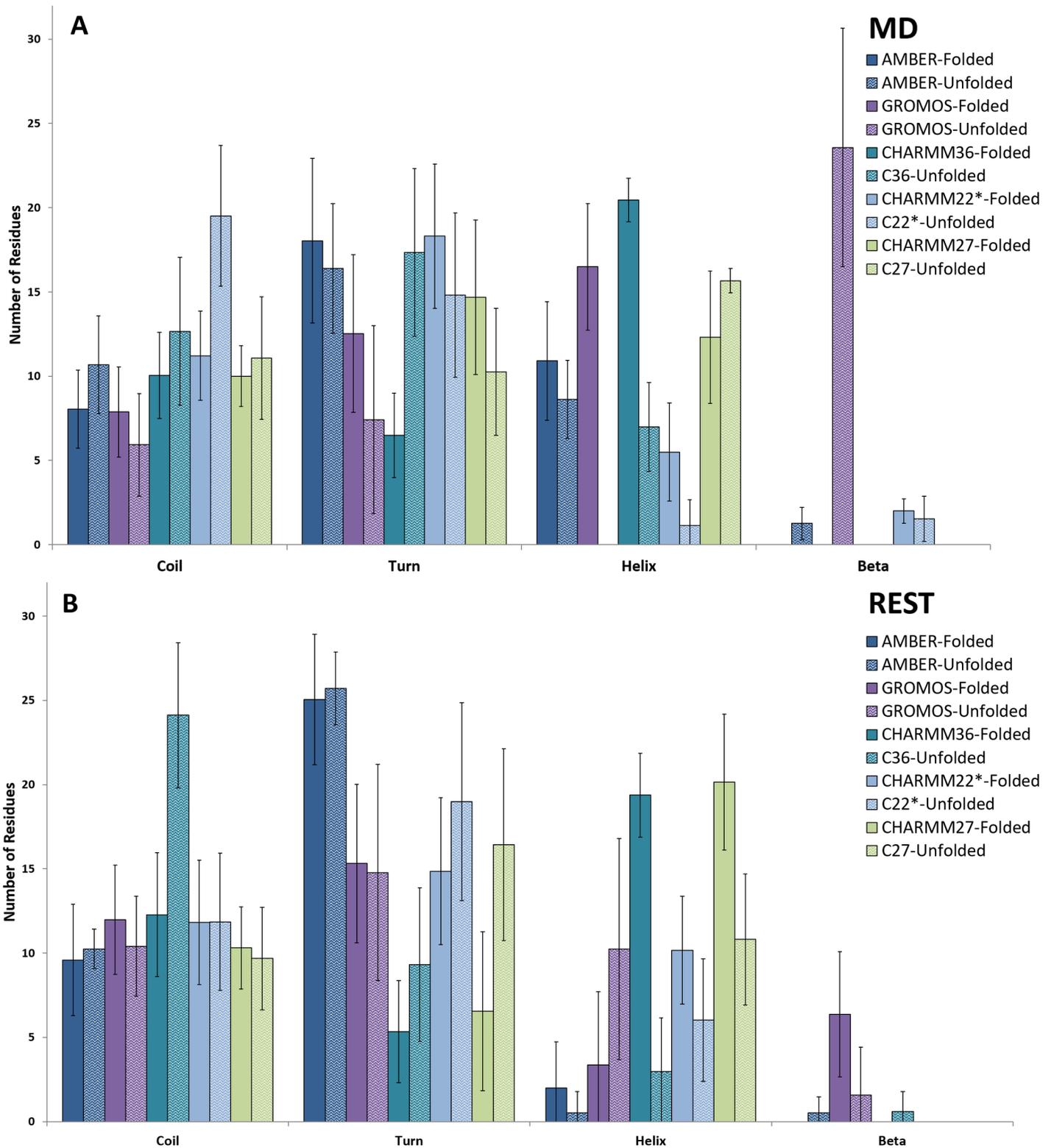


Fig 3. Average number of residues showing each secondary structure element determined on the equilibrated period of the (A) MD and (B) REST2 simulations for each forcefield and starting structure. The pattern filled bars are representing the unfolded runs while the solid filled bars with the same colour are the folded runs.

<https://doi.org/10.1371/journal.pone.0186219.g003>

conformation out of the low-energy wells, as seen from our simulations. Clearly, the starting structure influenced the resulting conformations despite the extensive conformational sampling. However, the purpose of this study was to investigate the ability of common forcefields to take the protein out of the helical conformation into the disordered solution state as seen in experimentally. We have shown that REST2 is able to mitigate some of the conformational biases exhibited by the forcefields, provided that the starting structure is not already folded in line with the forcefield's own preferences.

Comparison to other forcefield and enhanced-sampling studies. In comparison to the REMD and BEMD results from Zerze et al. using the Amberff03w forcefield, our simulations using REST2 with AMBER99SB*-ILDN yielded slight variations in secondary structures obtained. Whilst these two AMBER forcefields are very different from one another, our results demonstrate that drastic differences persist within different versions of the same family forcefield. AMBER99SB*-ILDN, originating from the ff99SB and ff03 forcefields, have a combination of improved side-chain torsion potentials for four amino acids (-ILDN), and a modification to the backbone dihedral potentials (ff99SB*) to provide better energetic balance between helix and coil conformations. [31, 39] Amberff03w, on the other hand, was derived from Amberff03 with the TIP4P water model, with a small backbone modification to match the population of helical states to experiment. [40, 41] These differences in modifications to the forcefield partly contribute to the variance of sampled conformations between these studies. Amberff03w in REMD and BEMD simulations produced more helical and coil content (13–15 residues as opposed to 10 using REST2, see supporting information S1 Fig; however far less turns (13 residues as opposed to 25 residues seen in our study). Amberff03w also featured β -structure formation (~2 residues) consistent between REMD and BEMD. This was not the case for our REST2 and MD simulations with AMBER99SB*-ILDN where REST2 resulted in an overall random coil conformation, whilst some helical content (~7 residues) remained within the timeframe used in the brute-force MD simulation. Since the Amberff03w forcefield was derived from Amberff03 with the TIP4P water model, this forcefield was found to produce more cooperative helix-coil transitions relative to the forcefield it was based on. [41] Nonetheless, the observed differences in sampled structures further illustrate the importance of using advanced sampling methods when modelling IDPs. On the other hand, the AMBER99SB forcefield used by Qiao et al. [16] obtained similar secondary structures to our AMBER99SB*-ILDN results. The secondary structures formed in their REMD study showed that approximately 16 residues were in a random coil conformation, 15 residues in bend or turn conformation, 3 in helical conformations and 3 in β -conformation. This is fairly consistent with the results we obtained in our AMBER99SB*-ILDN forcefield study. In a different forcefield study, Dupuis et al. found that the AMBER ff96 forcefield and REMD of human amylin with +3 and +4 protonated states formed significantly more β -sheets than other studies using only the +2 protonated state. [10] Their results showed approximately 10–13 residues in β -sheet conformations for +3 and +4 hIAPP respectively. Only a small portion of residues were in helical conformations, (~7 residues), with the rest making up turns and coiled conformations. Therefore, in addition to forcefield and water models, the protonation state of the protein also plays a significant role in the ambient conformation of a protein. For the purposes of this study, we had chosen the +2 protonation state.

In two separate studies, Hoffman et al. and Zerze et al. utilised the Amberff03w forcefield with the BEMD method to investigate the conformational behaviour of amylin. Their secondary structure results showed minor differences in helical versus strand content despite using the same technique and forcefield (Supporting Information S1 Fig). This variation demonstrates the importance of selection and definition of appropriate collective variables in BEMD simulations, where biases towards one or another secondary structure can play an important

role in the conformations formed and sampled. It is also worth noting that Hoffman et al. used different water models for their choice of forcefields. Specifically, the TIP4P[42] model was used for the Amberff03w and CHARMM22* simulations, while TIP3P[43] and SPC[44] water models were used for AMBER99SB*-ILDN and GROMOS96 53a6 respectively. The TIP4P water model was proposed in 2005 as a rigid four site model consisting of three fixed point charges and a single LJ centre with a negatively charged dummy atom along the HOH bisector. While this improves the electrostatic description of the molecule, it renders the model less computationally efficient relative to TIP3P.[42, 43] Recently, an updated version of TIP4P was developed by Piana et al.[45] and its effectiveness on IDPs were tested on Histatin 5 using the AMBER99S-B-ILDN forcefield. [46] Here, to investigate the role of the water model on the conformational preferences of folded amylin, brute-force MD and REST2 simulations were repeated using the CHARMM22* and CHARMM27 forcefields and the modified TIP3P (TIP3SP) water model (see Supporting Information S2 Fig). The secondary structure analysis showed the protein had slightly more helical content in the CHARMM27/TIP3SP MD and REST2 simulations, as well as the CHARMM22* REST2 simulations, compared to the simulations using the standard TIP3P water model. However, the TIP3SP CHARMM22* MD simulations, had no observed helical formations, as opposed to the ~5 residue helix seen in standard TIP3P. Nevertheless, the results demonstrated no significant differences between the structures obtained, apart from the small increase in helical content observed in the REST2 simulations. In a separate study, Boonstra et al. [47] suggested the use of unmodified TIP3P water model for modelling disordered proteins, due to a) the minor structural differences observed between the two water models for an intrinsically disordered protein and b) the reduction in computational cost. Following this, the unmodified TIP3P was chosen for testing water models in this study. Furthermore, as Henriques et al. [46] highlighted an accurate protein-water interaction was important for modelling unfolded IDPs and possibly the discrepancies between previous studies [4, 5] and our results would be attributed to the differences between water models used. In addition, a more recent CHARMM36m forcefield was proposed by Huang et al. [48] which improve upon the accuracy in generating backbone conformational ensembles for intrinsically disordered peptides and proteins. It is important to point that this study postulated that there is no single universal parameter applicable to all IDPs, and that IDP compatible water models may provide more accurate results.

Comparison between REST2 and BEMD

Numerous studies in the past had investigated the conformation of monomeric human amylin using various computational and experimental techniques.[1, 4, 5, 9–11, 49, 50] REMD and BEMD techniques previously employed by Zerze et al. and Hoffman et al. are good examples of applying advanced sampling methods to overcome the known limitations of spontaneous MD simulations, especially when dealing with IDPs. REST2 simulations from this study provide additional quantitative comparison of the advanced sampling approaches for modelling IDPs.

Metadynamics based approaches involve Gaussian potentials added to the forcefield to adaptively bias the MD simulation along predetermined reaction coordinates. These reaction coordinates are classified as the Collective Variables (CVs), typically including α -RMSD, β -RMSD, number of hydrogen bonds, or the number of atomic contacts in a biomolecule, to name a few. Bias-exchange metadynamics accelerates the sampling of metadynamics by using a series of replicas under the control of different biasing potentials and adapting them as the simulation proceeds.[4] On the other hand, REST2 achieves efficient sampling by applying a scaling factor to the potential energy of each replica, as well as reducing the amount of solvent sampling by only tempering the solute of the system.

It has been argued that the choice of CVs is the most important factor in determining the convergence and efficiency of the free-energy calculations in BEMD. [4, 5] Thus, if the chosen CVs are unable to capture the slowest degree of freedom to allow for conformational transitions, or are unable to distinguish between different metastable states, not all regions of the conformational space will be explored. This phenomenon of BEMD simulations is known as hysteresis.[4] Therefore, one would need some preliminary knowledge of the topological, chemical and physical properties of the system. This can, in principle, be overcome by using a very large group of CVs, effectively reducing the importance of selecting the “appropriate” CVs, although this will significantly increase the computational cost.[4] In contrast, in REST2 simulations, the most important factors are the number of replicas used and the highest effective temperature of the solute. Interestingly, a majority of the secondary structures sampled by our REST2 simulations agreed with the results of Zerze et al. and Hoffman et al., with the exception to CHARMM22* and GROMOS96 54a7, although Hoffman et al. used a slightly older variant of the GROMOS forcefield. The 54a7 version of the GROMOS forcefield used here incorporates a new helical propensity adjustment via the torsion term. This may be attributed to the higher helical content in monomeric amylin observed in our simulations starting from the folded amylin structure, compared to an overall β -hairpin structure sampled with 53a6 forcefield.[5] Furthermore, it is important to point out that the studies reviewed here have not considered different starting structures for their simulations, thus any dependence or bias from the initial conformation was not explored. Our REST2 simulations results clearly demonstrated that starting from unfolded conformation REST2 was able to alleviate some of the forcefield biases associated with CHARMM27 and GROMOS96 54a7 forcefields.

To complement the analyses of the sampling abilities of the techniques and the forcefield biases, the final point of comparison is the computational cost of each methodology. T-REMD and BEMD simulations by Zerze et al. were run for 8 μ s and 2.4 μ s each or 200 ns and 600 ns per replica respectively (total of 40 replicas). On the other hand, Hoffman et al. ran their BEMD simulations for 600 ns for all forcefields except AMBER99SB*-ILDN which was ran for 1 μ s. In contrast, the folded REST2 simulations in this work were ran for 100 ns per replica, with the exception to GROMOS96 54a7 that ran for 200 ns per replica; effectively resulting in a simulation time of 1.6 μ s and 3.2 μ s respectively. Conversely, the REST2 simulations starting from the unfolded state ran for 30 ns per replica, resulted in an effective sampling time of 480 ns, closely matching the simulated time of the respective MD runs. The results from this study indicated that for some forcefield REST2 simulations converged to different structures when performed from the unfolded state. Therefore, REST2 remains more efficient relative to other sampling techniques due to the demonstrated improvements in sampling conformational space in conjunction with potentially biasing forcefields.

Conclusion

In order to appropriately sample the free-energy landscape of an intrinsically disordered protein, one must perform a systematic analysis utilising the capabilities of advanced sampling techniques and a forcefield capable of reflecting the conformational preferences of such a protein. [3] REST2 simulations were benchmarked against spontaneous MD, T-REMD and BEMD approaches to highlight the advantages of REST2 relative to the other sampling methods. As expected, REST2 simulations sampled a much broader conformational space than brute force MD simulations while yielding agreeable results with MD, T-REMD and BEMD simulations across most forcefields. However, the conformational biases exhibited by some commonly used forcefields appear to persist in simulations by brute-force MD and even by advanced sampling technique such as REST2. By utilising REST2 in a systematic approach

using different starting structures we demonstrated that it is possible to improve the sampling of the conformational landscape of an intrinsically disordered protein such as amylin. Starting from the folded and unfolded conformations, the CHARMM27 forcefield favoured helical conformations, whereas CHARMM36 only sampled helical conformations when simulations were initiated from the folded structure. The GROMOS96 54a7 forcefield caused amylin to form some β -hairpin structures as observed in previous studies [5] and the AMBER99SB*-ILDN forcefield, resulted in a fully disordered conformation. Finally, CHARMM22* produced results in best agreement with experimental data, [12–14] providing a balance between disordered and helical conformations for amylin in solution, as experimental results indicated. [12] Overall, CHARMM22* is shown to be less dependent on the starting structure or sampling method in its ability to reproduce experimentally observed solution structures of amylin. The observed agreement with previous computational studies and experiments highlighted the ability of REST2 to explore the wider conformational space of amylin efficiently and effectively, compared to brute-force MD. However, it is important to note that qualitative experimental consistency of the simulation results from advanced sampling methods such as REST2 can be affected by the chosen forcefield, and no one forcefield should be considered universal for all IDPs.

Supporting information

S1 Fig. Average number of residues showing each secondary structure element. This is determined on the equilibrated period of the folded REST2 simulations for each forcefield. §Data taken from Zerze et al. (4) §§Data taken from Hoffman et al (5).

(DOCX)

S2 Fig. Secondary structure calculations for (A) MD and (B) REST2 simulation runs conducted with the modified TIP3SP water model. The average number of residues showing each secondary structure element determined over the equilibrated period.

(DOCX)

S3 Fig. Free energy calculations for human amylin. This was illustrated as a function of α and β RMSD for each of the forcefields investigated in this study starting from the folded conformation. REST2 free-energy plots are on the left-hand side, and MD free-energy plots are on the right hand side.

(DOCX)

S1 Table. Clustering structures from TIP3SP simulations. The most favourable conformations and their population determined from clustering analysis over the last 50 ns (MD) and last 5 ns (REST2) of the CHARMM22* and CHARMM27 simulations with the TIP3SP water model. The protein secondary structure is represented as cartoon with the α -helix coloured in purple, 3-10-helix in blue, turn in cyan and coil in white and extended β -sheet in yellow.

(DOCX)

Acknowledgments

The authors acknowledge the generous allocation of high performance computational resources from the Australian National Computational Infrastructure (NCI), the Pawsey Supercomputing Centre, and Melbourne Bioinformatics. Special thanks to Dr. Adam Makarucha for his contributions to the implementation of REST2 and general guidance throughout the early stages of this research.

Author Contributions

Conceptualization: Irene Yarovsky.

Data curation: Nevena Todorova.

Formal analysis: Enxi Peng, Nevena Todorova.

Funding acquisition: Irene Yarovsky.

Investigation: Enxi Peng, Nevena Todorova, Irene Yarovsky.

Methodology: Nevena Todorova, Irene Yarovsky.

Project administration: Irene Yarovsky.

Resources: Irene Yarovsky.

Supervision: Nevena Todorova, Irene Yarovsky.

Validation: Enxi Peng, Nevena Todorova.

Visualization: Nevena Todorova.

Writing – original draft: Enxi Peng.

Writing – review & editing: Nevena Todorova, Irene Yarovsky.

References

- Morriss-Andrews A, Shea J-E. Simulations of Protein Aggregation: Insights from Atomistic and Coarse-Grained Models. *The Journal of Physical Chemistry Letters*. 2014; 5(11):1899–908. <https://doi.org/10.1021/jz5006847> PMID: 26273871
- Latour RA. Perspectives on the simulation of protein-surface interactions using empirical force field methods. *Colloids Surf B: Biointerfaces*. 2014; 124:25–37. Epub 2014 Jun 30. doi: [10.1016/j.colsurfb.2014.06.050](https://doi.org/10.1016/j.colsurfb.2014.06.050). WOS:000347596200004; PubMed Central PMCID: PMC4252847. PMID: 25028242
- Todorova N, Legge FS, Treutlein H, Yarovsky I. Systematic comparison of empirical forcefields for molecular dynamic simulation of insulin. *Journal of Physical Chemistry B*. 2008; 112(35):11137–46. Epub 2008 Aug 12. <https://doi.org/10.1021/jp076825d> PMID: 18698702
- Zerze GH, Miller CM, Granata D, Mittal J. Free Energy Surface of an Intrinsically Disordered Protein: Comparison between Temperature Replica Exchange Molecular Dynamics and Bias-Exchange Metadynamics. *Journal of chemical theory and computation*. 2015; 11(6):2776–82. <https://doi.org/10.1021/acs.jctc.5b00047> PMID: 26575570
- Hoffmann KQ, McGovern M, Chiu C-C, de Pablo JJ. Secondary Structure of Rat and Human Amylin across Force Fields. *PloS one*. 2015; 10(7):e0134091. doi: [10.1371/journal.pone.0134091](https://doi.org/10.1371/journal.pone.0134091). PubMed Central PMCID: PMC4519342. PMID: 26221949
- Leach A. *Molecular Modelling: Principles and Applications (2nd Edition)*: Prentice Hall; 2001.
- Henriques J, Cragnell C, Skepö M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of chemical theory and computation*. 2015; 11(7):3420–31. <https://doi.org/10.1021/ct501178z> PMID: 26575776
- Shaw DE, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*. 2008; 51(7):91. <https://doi.org/10.1145/1364782.1364802>
- Reddy AS, Wang L, Singh S, Ling YL, Buchanan L, Zanni MT, et al. Stable and metastable states of human amylin in solution. *Biophysical Journal*. 2010; 99(7):2208–16. doi: [10.1016/j.bpj.2010.07.014](https://doi.org/10.1016/j.bpj.2010.07.014). PubMed Central PMCID: PMC3042569. PMID: 20923655
- Dupuis NF, Wu C, Shea J-E, Bowers MT. Human Islet Amyloid Polypeptide Monomers Form Ordered β -hairpins: A Possible Direct Amyloidogenic Precursor. *Journal of the American Chemical Society*. 2009; 131(51):18283–92. <https://doi.org/10.1021/ja903814q> PMID: 19950949
- Chun Wu J-ES. Structural Similarities and Differences between Amyloidogenic and Non-Amyloidogenic Islet Amyloid Polypeptide (IAPP). *PLOS Comput Biol*. 2013; 9(8):e1003211. <https://doi.org/10.1371/journal.pcbi.1003211> PMID: 24009497

12. Goldsbury C, Goldie K, Pellaud J, Seelig J, Frey P, Muller SA, et al. Amyloid fibril formation from full-length and fragments of amylin. *J Struct Biol.* 2000; 130(2–3):352–62. <https://doi.org/10.1006/jsbi.2000.4268> PMID: 10940238
13. Yonemoto IT, Kroon GJA, Dyson HJ, Balch WE, Kelly JW. Amylin proprotein processing generates progressively more amyloidogenic peptides that initially sample the helical state. *Biochemistry.* 2008; 47(37):9900–10. Epub 11. <https://doi.org/10.1021/bi800828u> PMID: 18710262
14. Cort JR, Liu Z, Lee GM, Huggins KN, Janes S, Prickett K, et al. Solution state structures of human pancreatic amylin and pramlintide. *Protein engineering, design & selection: PEDS.* 2009; 22(8):497–513. <https://doi.org/10.1093/protein/gzp029> PMID: 19596697; PubMed Central PMCID: PMC2719500.
15. Nanga RP, Brender JR, Vivekanandan S, Ramamoorthy A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica et biophysica acta.* 2011; 1808(10):2337–42. Epub 2011 Jun 23. doi: 10.1016/j.bbame.2011.06.012. WOS:000294982000001; PubMed Central PMCID: PMC3156962. PMID: 21723249
16. Qiao Q, Bowman GR, Huang X. Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation. *J Am Chem Soc.* 2013; 135(43):16092–101. <https://doi.org/10.1021/ja403147m> PMID: 24021023
17. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett.* 1999; 314(1–2):141–51. Epub 1999 Dec 20. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9)
18. Wright LB, Walsh TR. Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: insights from a quartz binding peptide. *Phys Chem Chem Phys.* 2013; 15(13):4715–26. doi: 10.1039/c3cp42921k. PubMed Central PMCID: PMC23423310. PMID: 23423310
19. Bussi G, Gervasio FL, Laio A, Parrinello M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society.* 2006; 128(41):13435–41. <https://doi.org/10.1021/ja062463w> PMID: 17031956
20. Terakawa T, Kameda T, Takada S. On easy implementation of a variant of the replica exchange with solute tempering in GROMACS. *Journal of Computational Chemistry.* 2011; 32(7):1228–34. Epub 2010 Nov 29. <https://doi.org/10.1002/jcc.21703> PMID: 21425280
21. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of chemical physics.* 1953; 21(6):1087–92. <https://doi.org/10.1063/1.1699114>
22. Todorova N, Marinelli F, Piana S, Yarovsky I. Exploring the folding free energy landscape of insulin using bias exchange metadynamics. *Journal of Physical Chemistry B.* 2009; 113(11):3556–64. Epub 2009 Feb 25. <https://doi.org/10.1021/jp809776v> PMID: 19243106
23. Wang L, Friesner RA, Berne BJ. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *Journal of Physical Chemistry B.* 2011; 115(30):9431–8. Epub 2011 Jun 29. doi: 10.1021/jp204407d. PubMed Central PMCID: PMC3172817. PMID: 21714551
24. Liu P, Kim B, Friesner RA, Berne BJ. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *PNAS.* 2005; 102(39):13749–54. Epub 2005 Sep 19. doi: 10.1073/pnas.0506346102. PubMed Central PMCID: PMC1236566. PMID: 16172406
25. Lundborg M, Lindahl E. Automatic GROMACS topology generation and comparisons of force fields for solvation free energy calculations. *Journal of Physical Chemistry B.* 2015; 119(3):810–23. <https://doi.org/10.1021/jp505332p> PMID: 25343332
26. Darden T, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys.* 1993; 98(12):10089–92. <https://doi.org/10.1063/1.464397>
27. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys.* 1984; 81(8):3684–90. <https://doi.org/10.1063/1.448118>
28. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007; 126(1):014101. <https://doi.org/10.1063/1.2408420> PMID: 17212484
29. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry.* 1997; 18(12):1463–72. Epub 1998 Dec 7. [https://doi.org/10.1002/\(Sici\)1096-987x\(199709\)18:12<1463::Aid-Jcc4>3.0.Co;2-H](https://doi.org/10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H)
30. Piana S, Lindorff-Larsen K, Shaw David E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophysical journal.* 2011; 100(9):L47–L9. Epub 2011 May 3. <https://doi.org/10.1016/j.bpj.2011.03.051> PMID: 21539772
31. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct Funct Bioinf.* 2010; 78(8):1950–8. Epub 210 Mar 9. doi: 10.1002/prot.22711. PubMed Central PMCID: PMC2970904. PMID: 20408171

32. Huang W, Lin Z, van Gunsteren WF. Validation of the GROMOS 54A7 Force Field with Respect to β -Peptide Folding. *Journal of chemical theory and computation*. 2011; 7(5):1237–43. Epub 2011 Apr 20. <https://doi.org/10.1021/ct100747y> PMID: 26610119
33. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry*. 2013; 34(25):2135–45. Epub 2013 Jul 6. <https://doi.org/10.1002/jcc.23354> PMID: 23832629
34. Mackerell AD, Jr., Feig M, Brooks CL, III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*. 2004; 25(11):1400–15. doi: [10.1002/jcc.20065](https://doi.org/10.1002/jcc.20065). PubMed Central PMCID: PMC15185334. PMID: 15185334
35. Palafox-Hernandez JP, Tang Z, Hughes ZE, Li Y, Swihart MT, Prasad PN, et al. Comparative study of materials-binding peptide interactions with gold and silver surfaces and nanostructures: A thermodynamic basis for biological selectivity of inorganic materials. *Chemistry of Materials*. 2014; 26(17):4960–9. Epub 2014 Aug 29. <https://doi.org/10.1021/cm501529u>
36. Shao J, Tanner SW, Thompson N, Cheatham TE, III. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of chemical theory and computation*. 2007; 3(6):2312–34. <https://doi.org/10.1021/ct700119m> PMID: 26636222
37. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; 14(1):33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
38. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*. 2009; 180(10):1961–72. <https://doi.org/10.1016/j.cpc.2009.05.011>
39. Best RB, Hummer G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The journal of physical chemistry B*. 2009; 113(26):9004–15. <https://doi.org/10.1021/jp901540t> PMID: 19514729; PubMed Central PMCID: PMC3115786.
40. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic validation of protein force fields against experimental data. *PloS one*. 2012; 7(2):e32131. <https://doi.org/10.1371/journal.pone.0032131> PMID: 22384157; PubMed Central PMCID: PMC3285199.
41. Best RB, Mittal J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *Journal of Physical Chemistry B*. 2010; 114(46):14916–23. <https://doi.org/10.1021/jp108618d> PMID: 21038907
42. Abascal JL, Vega C. A general purpose model for the condensed phases of water: TIP4P/2005. *J Chem Phys*. 2005; 123(23):234505. <https://doi.org/10.1063/1.2121687> PMID: 16392929
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983; 79(2):926–35. <https://doi.org/10.1063/1.445869>
44. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. *Intermolecular Forces. The Jerusalem Symposia on Quantum Chemistry and Biochemistry*. Dordrecht, Netherlands: Springer; 1981. p. 331–42.
45. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *Journal of Physical Chemistry B*. 2015; 119(16):5113–23. <https://doi.org/10.1021/jp508971m> PMID: 25764013
46. Henriques J, Skepo M. Molecular dynamics simulations of intrinsically disordered proteins: on the accuracy of the TIP4P-D water model and the representativeness of protein disorder models. *Journal of chemical theory and computation*. 2016; 12(7):3407–15. <https://doi.org/10.1021/acs.jctc.6b00429> PMID: 27243806
47. Boonstra S, Onck PR, Giessen E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *The journal of physical chemistry B*. 2016; 120(15):3692–8. <https://doi.org/10.1021/acs.jpcc.6b01316> PMID: 27031562.
48. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017; 14(1):71–3. <https://doi.org/10.1038/nmeth.4067> PMID: 27819658; PubMed Central PMCID: PMC5199616.
49. Andrews MN, Winter R. Comparing the structural properties of human and rat islet amyloid polypeptide by MD computer simulations. *Biophys Chem*. 2011; 156(1):43–50. <https://doi.org/10.1016/j.bpc.2010.12.007> PMID: 21266296
50. Liang G, Zhao J, Yu X, Zheng J. Comparative molecular dynamics study of human islet amyloid polypeptide (IAPP) and rat IAPP oligomers. *Biochemistry*. 2013; 52(6):1089–100. <https://doi.org/10.1021/bi301525e> PMID: 23331123.