Benchmarking physical models of computer simulations for structural characterization across amyloidogenic IDPs

Shayan Bhattacharya and Damien Thompson
Department of Physics, Bernal Institute, University of Limerick, V94 T9PX, Ireland
shayon.bhattacharya@ul.ie
Keywords: Chemistry; Physics and Energy

Abstract
Conformational switches between monomers of amyloidogenic intrinsically disordered proteins (IDPs) such as amyloid β, α-synuclein, prion and tau might potentially act as the most fundamental trigger for the pathogenesis of neurodegenerative disorders like Alzheimer’s, Parkinson’s, transmissible spongiform encephalopathies etc. A group of IDPs herein referred to as ‘amyloidogenic IDPs’ leads to the formation of amyloids in different regions of the body.[2] Amyloids are protein aggregates folded into a certain shape, which facilitates protein aggregation, giving rise to large fibrillary structures. Amyloid formation in the brain is associated with several lethal diseases including Alzheimer’s, Parkinson’s, ALS, transmissible spongiform encephalopathies etc. to name but a few. The primary process of amyloid aggregation initiates by misfolding of monomers (IDPs) leading to aggregation into oligomers, and finally fibrils. Recent research suggests that small soluble oligomers are the most toxic entities in the pathway.[3] However, conformational switches within the amyloidogenic IDP monomers might sample distinct conformational signatures acting as the most fundamental trigger for amyloidogenesis.[4] Hence, structural characterization at the level of monomers to accurately account for the ensemble features remain extremely important.

Traditional bulk in vitro experimental techniques such as NMR are insufficient to account for such fast conformational changes of IDPs due to limitations in their sampling timescales. Hence, thorough sampling of configurational space is required through Boltzmann weighting implemented with molecular dynamics (MD) computer simulations as an essential means to study the behaviour of these proteins. In addition to ensuring sufficient sampling of phase space (important events occur inside the body at µs to ms time scales), another consideration is the accuracy of the physical models (such as the force fields and water models) used in simulating IDPs. The popular state-of-the-art protein force fields and water models have been shown to be highly incompetent, as they tend to overly collapse the IDPs and/or sample spurious secondary structures, when compared to experimental measurements such as NMR, small-angle x-ray scattering (SAXS), circular dichroism (CD) data etc.[5] Two recent developments in the force field department have shown that scaling the protein-water (Amber ff03ws)[5] or water-water (TIP4P-D)[6] nonbonded interaction parameters can accurately reproduce the experimental measures for some model IDPs through simulations.[7] In this work, we have used µs-scale MD simulations in explicit solvent to compare eight different combinations of protein and water force fields including the latest Amber ff03ws (A03WS) and TIP4P-D (D) across four amyloidogenic IDPs: amyloid β42 (Aβ42), α-synuclein (α-syn), cellular prion protein (PrP) and tau. Our results indicate that both A03WS and D perform the best, when it comes to modelling accuracy and benchmarking against experimental data.

1. Introduction
At least one-third of the human proteome is known to belong to the class of intrinsically disordered proteins (IDPs), or have intrinsically disordered regions (IDRs).[1] The native state of IDPs at near-physiological condition present a complex ensemble of fluctuating structures unlike folded proteins. This allows IDPs to be highly promiscuous with their interactions and functionally non-discrete. A group of IDPs herein referred to as ‘amyloidogenic IDPs’ leads to the formation of amyloids in different regions of the body.[2] Amyloids are protein aggregates folded into a certain shape, which facilitates protein aggregation, giving rise to large fibrillary structures. Amyloid formation in the brain is associated with several lethal diseases including Alzheimer’s, Parkinson’s, ALS,
Aβ42, 1XQ8 for α-syn, 2LSB for PrPc and 2MZ7 for tau. The peptides were appropriately solvated in water boxes adding physiological concentration of ions (0.15 m/L NaCl). The structures were minimized and equilibrated at 310K temperature and 1 bar pressure in both NVT and NPT conditions. Production run was performed with both NAMD and Gromacs MD engines for 2µs for Aβ42 and 1µs for all the other IDPs, amounting to a total run-time of ~45 µs, and generating data as big as 7TB of disk space. Eight different semi-empirically chosen combinations of protein force fields such as Charmm22* (C22*), Charmm36 (C36), Amber ff03 (A03) and Amber ff03ws (A03WS), and water models such as TIP3P, TIP4P, TIP4P-Ew (EW), TIP4P/2005 (2005) and TIP4P-D (D) were compared. The last 0.5µs or the last 1µs of the simulation trajectory was used for analysis, wherever applicable.

3. Analyses
We performed extensive analyses of simulation trajectories involving several different criteria that aids in accurately describing the IDP structural ensemble features while evaluating the effect of force fields, such as chain dimensions through radius of gyration, hydrogen bonds, conformational entropies, solvation energies, and back-calculation of experimental observables, such as NMR chemical shifts, J-couplings, residual dipolar couplings, SAXS intensity profiles etc.

4. Results and conclusion
From our analyses, we observe that the physical models which give larger weights to protein-water interactions, such as the protein and water force fields, Amber03ws and TIP4P-D, respectively, give significantly improved performance in terms of dynamical features, as well as correlation with experiments. This may provide a potentially predictive model for properties of yet poorly characterised amyloidogenic IDPs.

5. Figures and graphs

![Figure 1](image1.png)

**Figure 1.** Simulation-generated structural ensembles of Aβ42 by combinations of different force fields/water models.

![Figure 2](image2.png)

**Figure 2.** Examples of (A) histogram of radius of gyration, (B) back-calculated SAXS intensity curve, (C) population of secondary structures, and (D) back-calculated chemical shift values for different IDP systems.

8. References