Current Topics, Methods, and Challenges in the Modelling of Intrinsically Disordered Protein Dynamics

Rickie Xian; r4xian@uwaterloo.ca; MSc Candidate; Dr. Sarah Rauscher

Introduction

The paradigm that the primary amino acid sequence prescribes structure and thus function has for a long time been central to the understanding of Protein Science. Though the theory is supported by the behaviour of most structured proteins, it loses much of its applicability when discussing Intrinsically Disordered Proteins. These peculiar proteins, whose tertiary structure constantly interconverts between a series of energetically favourable conformations, are the root of many current, pressing scientific questions. Many biological processes that are still yet to be elucidated—the mechanisms of protein folding, ligand binding, and general protein dynamics—involves IDPs. Because most dynamic protein events are on such short time scales, using experimental methods to observe their action often times doesn’t yield useful data. As well, the data resulting from scientific techniques developed for structured, “static proteins” must be presented in conjunction with data from methods tailored specifically to IDPs in order to have significance.

A method that models IDPs with shocking accuracy is computer simulations, particularly Molecular Dynamics (MD) simulations. With computational power only recently increasing enough to encompass the timescale needed for protein dynamics, MD simulations are still fairly novel in their implementations. This paper will discuss and consolidate the current methods, problems, and solutions of using MD Simulation to model IDPs. Which simulation parameters can be altered to more precisely describe observed biological behaviour? How can one accurately use MD simulation to answer questions that, when using experimental methods, have no answer? How can the data resulting from MD simulation be analyzed and quantified to support the conclusions being drawn?
Force Fields: How Can MD Simulations be Tailored to IDPs?

It’s not surprising that many current improvements to IDP modelling with MD simulation have been made to the force field: the potential energy function applied during simulation. The specific forces of potential energy act on the system to inform its movement, thereby being essential to simulation dynamics. A common technique used by researchers to increase modelling accuracy is the altering of force field parameters to more closely match experimental data or analytical models.

The torsional potential which dictates the dihedral angles of both the backbone and side chain is an area of particular interest for force field improvement (13). Torsion potentials influence systems who exist at the interface of bonding and non-bonding interactions, thus these potentials are often modified when one seeks to reform non-bonded, long-range interactions (13). Attempting to more accurately develop these torsional parameters, researchers have included more subtle physical effects in their torsion potentials. For example, solvation effects weren’t previously considered when deriving torsion angle parameters, an approach that worked for shorter simulations of systems whose structures were well-determined, experimentally (7). However, Zgarbová et.al recognized that for longer time scales or more uncommon structures, revision of torsion parameters was necessary for accurate modelling (7). Zgarbová’s team re-derived dihedrals using glycosidic torsion as the model system; instead of deriving these parameters in vacuo as usual, Zgarbová et.al found the dihedrals from the difference in Quantum Mechanical reaction fields and the Poisson-Boltzmann continuum solvation models (7). The advantage to reparameterization of torsional parameters in continuum solution outside of vacuum was the inclusion of two previously overlooked, conformation-dependent effects: solute-solvent polarization and solvation of the solute’s charge distributions (7).

To further refine dihedrals, researchers often compared the data of the same system under influence of multiple force fields, for example the four variants of the AMBER force field on quadruplex DNA, RNA, and Z-DNA used by Krepl et.al to reformulate chi
dihedral angles (9). In these nucleic acid simulations, the best AMBER parameters were confirmed with comparison to X-ray crystallographic data to reparametrize the chi dihedral for syn-conformation nucleotide containing DNA (9). However, the accuracy of the results for IDPs was compromised by the comparison of simulation data to crystallographic data, wherein disordered regions around phosphate groups decreased the resolution of crystal structures (9). Thus from the results of Krepl et. al and Zgarbová et. al, selecting which experimental data to use as comparison for force field refinement is still a point of contention, especially if the system in question has significant disordered regions. In deriving potentials, most researchers choose to fit their simulation data to either experimental data, like NMR or structure-database data (13), or analytical calculations, like Zgarbová’s Quantum Chemistry derivation of dihedrals (7). A noteworthy result of torsional potential refinement was the decreasing of potential energy away from local minima such that simulation of folded states and multiple temperature systems could be more accurate relative to experimental results of equilibrium protein folding (13).

Another method of torsional potential and implicitly polarized atomic charge reformation was that of the AMBER ff15ipq force field done by Debiec et.al (8). By using the IPolQ charge model, Debiec’s team considered both dipole-dipole interactions under external electrostatic field and the influence of the water model on the solute charge distribution to fit atomic charges of the ff15ipq force field (8). Their methodology sharply contrasted with previous formation of AMBER force fields, which used the REsP method to fit atomic charges (8). As well, Debiec et.al fit torsional parameters simultaneously to the most likely conformation of all residues they appeared in rather than the conventional method of fitting only to glycine and alanine (8). Finally, with the usage of ab initio calculations of two correlated charge sets to construct new force fields, Debiec et.al showed the new ff15ipq force field didn’t corrupt secondary structure of globular proteins nor salt-bridge propensities, and also reliably predicted conformation of folded proteins and IDP’s (8). Moreover, by once again comparing simulation data to NMR data, it was clear the force field had strong
agreement with J-coupling data and relaxation rates, a testament to the physical accuracy of the ff15ipq force field for both ordered and disordered proteins (8).

Though continual revision of torsional potentials supports the importance of these parameters to force field development, many times systematic errors in MD simulations can’t be resolved with only changing this one class of parameters (13). Adjusting other potential energy function parameters, such as the water model, is imperative to further improvement of force field accuracy, as exemplified by Wang et.al’s AMBER-FB15 force field development (2). By updating the TIP3P water model to TIP3P-5B and, like Zgarbová’s team, borrowing from an existing AMBER force field, AMBER94, Wang et. al, optimized the potential energy landscape of the force field (2). Like the findings of Nerenberg et. al, the optimized potential energy of AMBER-FB15 resulted in lower energy away from the minima (2). With the new force field AMBER-FB15 and the TIP3P-5B water model, Wang et. al showed through comparison to experimental thermodynamic equilibrium quantities that AMBER-FB15 provided improved accuracy for protein simulation, especially in systems of high conformational changes, fluctuations away from the energy minima, and temperature dependency (2). Consequently, the newly developed AMBER-FB15 would be a more suitable force field for IDPs, as the rapid, temperature-dependent conformation interconversion and various energy minima sampling of these proteins is an example of the system AMBER-FB15 was formulated to target.

Long-range and solute-solute interactions, especially those of polar nature, are also a class of parameters to be revised by many, as already seen with the work of Debiec et.al (8). Miller et.al exemplified that osmotic coefficients are of specific interest when refining these force field interactions (3). By measuring osmotic coefficients of various combinations of the TIP3P and TIP4P-Ew water models paired with four existing force fields, AMBER ff99SB-ILDN, CHARMM36, GROMOS54a7, and OPLS-AA, Miller’s team computed osmotic coefficients lower than those of experiment (3). This coefficient discrepancy implied that solute-solute interactions were generally overly favourable in simulation (3). By altering the Van der Waals interactions of specific solute-solute pairs,
Miller et.al used osmotic pressure optimized parameters of the four aforementioned force fields to produce conformational data of higher agreement with experimental data (3).

Another subfield of interest of long-range interactions is the modelling of lipid-lipid and lipid-environment interactions in biological membranes, for example the work of Poger et.al (1). The modification of these interactions generally involves calculation of certain membrane properties from simulation, for example membrane thickness, and then comparison to experimental data such as X-ray and neutron scattering (1). The final class of force field parameter adjustments to be mentioned are secondary structure propensities, like that of the work of Best et.al (4). Best’s team made amendments to the backbone energy of two force fields, ff03 and ff99SB, and showed their optimized force fields produced helix propensities more consistent with NMR experiments of short, folded proteins (4). It should be noted this particular force field development may have limited application to IDPs, as these proteins have few, short-lived regions of ordered secondary structures since they constantly diffuse through their conformational landscapes.

**How Can Existing MD Simulation Techniques be Used to Understand IDPs?**

Besides redefining force field parameters, researches have shown IDPs can be understood by analyzing the effect of existing parameters on the system, for example the internal friction of unfolded and disordered proteins found from dihedral relaxation and intrachain interactions by Zheng et. al (12). Zheng’s team knew internal friction of IDPs, specifically ACTR, couldn’t just be from solvent viscosity, but was also due to chain collapse (12). By performing various MD simulations of ACTR in unique solvent viscosities, Zheng’s team explored the effects of dihedral relaxation and solvent-dependent dihedral angle flips on internal friction (12). Zheng et. al found that IDP internal friction is directly proportional to chain compactness due to an increased dihedral flipping free energy barrier; as well, the attraction within peptide chains
decreases the surface area that can be accessed by the solvent (12). Thus in expanded chains, Zheng et. al showed the internal friction of IDPs was mostly due to dihedral angle flips, whereas in compact chains, the friction was due to intrachain attraction.

Further MD simulations have been run with desires to understand folding behaviour of various IDPs and IDP-involving complexes. The folding dynamics of IDP-involving protein complexes are of notable interest, especially when folding is triggered by some ligand binding, like that of the work of Piana et.al (15) and Turjanski et.al (11). Both groups focused on transition states—how exactly are these high energy states between folding stabilized? Piana et.al examined the binding of the pKID domain of the transcription factor CREB to the KIX domain of CREB’s co-activator CBP (15). Using data analysis, for example that of phi-angle data, Piana et.al found that the binding transition state was greatly stabilized by native-like interactions; as well most notably of all, the pKID domain of CREB only folded after co-activator binding, a result consistent with the induced fit model of enzyme-ligand binding (15). Because of the high degree of agreement between simulation data and NMR studies, Piana et.al concluded that disorder in protein-ligand binding is advantageous, for it lends “promiscuity” to a binding pathway such that a single protein can bind multiple targets (15).

The other team, Turjanski et.al, studied the folding dynamics of protein substrates in the GroEL cavity, a bacterial chaperonin cavity responsible for isolating mis-folded proteins to prevent aggregation upon refolding (11). Using the a99SB-disp force field, Turjanski’s team found the GroEL tetradecamer to be extremely dynamic under simulation, transitioning between various states; as well, in the GroEL cavity, strong interactions between the unfolded protein substrate and the disordered residues of GroEL’s C terminus tails were observed (11). This disordered interaction stabilized the unfolded state of the substrate by roughening the free energy surface, allowing for slower folding within the cavity (11). Thus, both Piana et.al and Turjanski et.al used MD simulation to understand how IDPs were advantageous to protein binding and folding mechanisms.
Before the structure-function protein paradigm can be completely nullified for IDPs, one should consider the works of both Sasmal et.al and Marsh et. al, whose efforts have been made to understand IDP function by distilling their structure. Both Sasmal et.al and Marsh et.al used MD simulation to discretize IDP structure, with the former focusing on an amyloid-beta ensemble (5) and the latter on the drk N-terminal SH3 domain (10). Sasmal et.al used a nitroxide paramagnetic spin label on amyloid-beta42 and compared it with another unlabelled, computationally generated structure of the same sequence to observe the effect of the spin label on the amyloid structure (5). By performing independent calculations of NMR observables of both labelled and unlabelled peptide sequences, Sasmal’s team showed there exists a significant difference in the structure of the two peptides: the perturbations caused by nitroxide labelling caused beta-hairpin formation at the C-terminus (5). Also, though the structural perturbation of IDPs due to paramagnetic tags is more than that of folded proteins, Sasmal et.al demonstrated Nuclear Overhauser effects (NOE) experiments could still be performed without label dictation of NMR relaxation mechanisms (5).

Marsh et.al also attempted to clarify IDP structure, wherein structural ensembles of the unfolded drk N-terminal SH3 domain were generated with ENSEMBLE under denaturing conditions (10). In contrast to their previous methods, Marsh’s team included significantly more experimental restraints and parameters to more accurately specify the ensembles generated, for example the inclusion of small angle X-ray scattering methods, nitroxide paramagnetic relaxation enhancements, and nuclear Overhauser effects (with the latter two also explored by Sasmal et.al) (10). With the implementation of a novel iterative conformation sampling method, Marsh et.al generated structural ensembles of the SH3 domain more consistent with experimental data: more compact ensembles possessing non-native alpha-helices and some non-native tertiary structure (10). The work of both Sasmal et.al and Marsh et.al demonstrates the continual desire to further clarify the structure of IDPs by using various ab initio calculation, inclusion of more parameters, and computational methods, all the while comparing the resulting structures to established experimental data to test accuracy.
How does MD simulation Reveal the Differences between IDPs and Ordered Proteins? A Comparison of Two Data Sets

Methods
After thorough discussion of methods researcher’s have used to modify force fields and use simulations to understand IDPs, it’d be useful to discuss how IDP MD simulation data differs from that of structured proteins. To explore this difference, two simulations on GROMACS version 2018.4 were carried out. The first simulation was a generic, folded protein simulation of lysozyme in water. The second was a high temperature simulation of a six residue fragment of the alpha-beta peptide, KLVFFA, wherein the peptide’s conformation sampling under high temperature emulated the rapid interconversion of IDP tertiary structure. The lysozyme simulation was run for 20.0 nanoseconds in a cubic box using the all-atom OPLS force field, whereas the KLVFFA simulation was run for 7.0 nanoseconds using the CHARMM36m force field in a rhombic dodecahedron box. It should be noted that the KLVFFA peptide was intended to be run for 20.0 nanoseconds, however due to failure to properly correct for periodic boundary conditions, the system became unstable after 7.0 nanoseconds, causing the production run to stop prematurely.

The same sequential methodology was applied to both simulation systems: topology generation from a PDB file, box definition and system solvation, energy minimization, temperature and pressure equilibration, and finally production run. It should be mentioned that an eight-peptide system was constructed fro the KLVFFA peptide wherein eight peptide conformations were randomly selected from unique time steps to represent eight conformations of the IDP. Should the reader desire further information on methods, a detailed methodology for both simulations can be found under supporting materials.
Results:
The energy minimization and pressure equilibration data of the lysozyme system are consistent with expected data of past simulations. It’s clear the lysozyme system minimizes its energy at around 6.25e5 kJ/mol, with its pressure equilibrating at around 0 bars, and its density quickly approaching, then equilibrating around 1020 kg/m^3. The minimization and equilibration data for the KLVFFA peptide was not included due to near universal similarities in the behaviour of these data regardless of system choice. The RMSD of the lysozyme simulation had near constant fluctuations about 0.1nm for all 20.0 nanoseconds, and its radius of gyration first fluctuated about 1.45 nm, then at around 5.0 nanoseconds, gently decreased to fluctuations about 1.4 nm.
The radius of gyration data for the KLVFFA peptide displays shocking periodicity and suspiciously large jumps of over 150 nm; the RMSD data for the peptide was nearly identical to the Rg data and thus not included. These data are completely unphysical artifacts of the failure to properly correct for periodic boundary conditions (PBCs), as mentioned above. When the peptide reached one end of the simulation box, instead of continuously moving through to the other end of the box, it “jumped” from one end to another, causing large RMSD jumps. It’s clear the Rg and RMSD data for KLVFFA isn’t physically significant, so a comparison between the two systems for Rg and RMSD trajectories of the production runs can’t be made.

Regardless of the failure to correct for PBCs, the total number of hydrogen bonds within the protein versus time can still be compared for both systems. With comparison of the two hydrogen bond data sets, it’s clear lysozyme displays fluctuation about 90 hydrogen bonds, while the KLVFFA system fluctuates first around 32.5 hydrogen bonds, then decreases to flux about 30 hydrogen bonds after 1.5 nanoseconds. The three fold increase in hydrogen bonds for lysozyme relative to that of the KLVFFA peptide is expected, as an ordered protein will have more intra-protein hydrogen bonds than that of a disordered protein with variant 3D orientation. Thus, the fluctuation of an IDP’s 3D structure is demonstrated by the more drastic fluctuation of the number of hydrogen bonds of that of the KLVFFA system relative to lysozyme’s near constant fluctuations about a central value. However, the true extent of KLVFFA h-bond number fluctuation requires a longer production run so as not to improperly extrapolate shorter simulation data for comparison with lysozyme’s 20.0 nanosecond production run.

**Conclusion:**

In conclusion, though the work presented exemplifies the immense efforts made to model and understand IDPs, more work is needed for more precise understanding. Further work should be done to optimize the parameters of MD simulation force fields, specifically those of torsional potentials, water models, and solute and solvent interactions, to more accurately match experimental data. The structure of IDPs also
needs further simulations to understand, especially under more physically relevant conditions, for example denaturing conditions, which more closely mimic biological environments. Because IDPs are still so elusive, the results and conclusions presented above shouldn’t be taken as absolute fact, but rather, with careful scrutiny. For though there is no universally agreed upon “best methodology” to understand IDPs, the aforementioned methods, if not unanimously supported, are nonetheless useful in exemplifying the various attempts made thus far to try to understand the intriguing nature of IDPs.

Supporting Materials:

1) Meneksedag-Erol, D., & Rauscher, S. (2018, August 17). Simulations of Protein Aggregation. Department of Chemical and Physical Sciences, University of Toronto Mississauga.

2) Lemkul, J. A. (2018). Lysozyme in Water. In GROMACS Tutorial. Retrieved from http://www.mdtutorials.com/gmx/lysozyme/index.html

References:

1) David Poger, Bertrand Caron, Alan E. Mark, Validating lipid force fields against experimental data: Progress, challenges and perspectives, BBA - Biomembranes (2016), doi: 10.1016/j.bbamem.2016.01.029

2) Wang, L., McKiernan, K. A., Gomes, J., Beauchamp, K. A., Head-Gordon, T., Rice, J. E., & Swope, W. C. (2017, March 17). Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. The Journal of Physical Chemistry. doi:10.1021/acs.jpcb.7b02320

3) J Chem Theory Comput. 2017 April 11; 13(4): 1812–1826. doi:10.1021/acs.jctc.6b01059.
4) Best, R. B., & Hummer, G. (2009, May 14). Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem.*, 113, 9004-9015.

5) Sasmal, S., Lincoff, J., & Head-Gordon, T. (2017, September 5). Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-b. *Biophysical Journal*, 113, 1002-1011.

6) Pestana, L. R., Minnetian, N., Lammers, L. N., & Head-Gordon, T. (2017, December 26). Dynamical inversion of the energy landscape promotes non-equilibrium self-assembly of binary mixtures. *Chem. Sci*, 9(1640). doi:10.1039/c7sc03524a

7) Zgarbová, M., Luque, F. J., Sponer, J., Otyepka, M., & Jurecka, P. (2012, August 21). A Novel Approach for Deriving Force Field Torsion Angle Parameters Accounting for Conformation-Dependent Solvation Effects. *Journal of Chemical Theory and Computation*, 8, 3232-3242. doi:10.1021/ct3001987

8) Debiec, K. T., Cerutti, D. S., Baker, L. R., Gronenborn, A. M., Case, D. A., & Chong, L. T. (2016, July 11). Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *Journal of Chemical Theory and Computation*, 12, 3926-3947. doi:10.1021/acs.jctc.6b00567

9) *J Chem Theory Comput.* 2012 July 10; 8(7): 2506–2520. doi:10.1021/ct300275s

10) Marsh, J. A., & Forman-Kay, J. D. (2009, June 9). Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J. Mol. Biol.*, 391, 359-374. doi:10.1016/j.jmb.2009.06.001
11) *J. Phys. Chem. B*, Just Accepted Manuscript: Atomic-Level Description of Protein Folding inside the GroEL Cavity • DOI: 10.1021/acs.jpcb.8b07366 • Publication Date (Web): 02 Oct 2018 Downloaded from http://pubs.acs.org on October 12, 2018

12) Zheng, W., Hofmann, H., Schuler, B., & Best, R. B. (2018, September 16). Origin of Internal Friction in Disordered Proteins Depends on Solvent Quality. *J. Phys. Chem.*, 391. doi:10.1021/acs.jpcb.8b07425

13) Nerenberg, P. S., & Head-Gordon, T. (2018, February 22). New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology*, 129-138. doi:10.1016/j.sbi.2018.02.002

14) Grossfield, A., Patrone, P. N., Roe, D. R., Schultz, A. J., Siderius, D. W., & Zuckerman, D. M. (2018, October 17). Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations. *Living J. Comp. Mol. Sci.*

15) Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-Induced Folding of a Natively Unstructured Transcription Factor. PLoS Comput Biol 4(4): e1000060. doi:10.1371/journal.pcbi.1000060