Progress in Biophysics and Molecular Biology xxx (2015) 1–6

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Progress in studying intrinsically disordered proteins with atomistic simulations

Nathaniel Stanley [a], Santiago Esteban-Martín [a, b], Gianni De Fabritiis [a, c, *]

[a] Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain
[b] Joint BSC-IRB-CRG Research Programme in Computational Biology, Barcelona Supercomputing Center – BSC, Jordi Girona 29, 08034 Barcelona, Spain
[c] Institució Catalana de Recerca i Estudis Avançats, Passeig Lluis Companys 23, 08010 Barcelona, Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Intrinsically disordered proteins
High-throughput molecular dynamics
Markov state models

ABSTRACT

Intrinsically disordered proteins are increasingly the focus of biological research since their significance was acknowledged over a decade ago. Due to their importance in biomolecular interactions, they are found to play key roles in many diseases such as cancers and amyloidoses. However, because they lack stable structure they pose a challenge for many experimental methods that are traditionally used to study proteins. Atomistic molecular dynamics simulations can help get around many of the problems faced by such methods provided appropriate timescales are sampled and underlying empirical force fields are applicable. This review presents recent works that highlight the power and potential of atomistic simulations to transform the investigatory pipeline by providing critical insights into the behavior and interactions of intrinsically disordered proteins.

Contents

1. Introduction

Intrinsically disordered proteins (IDPs) are proteins that lack fixed secondary and tertiary structure. This makes them particularly difficult to characterize by traditional biophysical techniques like X-ray crystallography, which in part led to their being largely ignored for many years. They finally gained widespread acceptance around the turn of the millennium as their prevalence and functional significance became clear (Dunker et al., 2001; Tompa, 2009, 2002; Uversky et al., 2000; Wright and Dyson, 1999).

Abbreviations: IDP, intrinsically disordered protein; MD, molecular dynamics; HTMD, high-throughput molecular dynamics; MSM, Markov state models; KID, kinase inducible domain; NMR, nuclear magnetic resonance; PRE, Paramagnetic relaxation enhancement; PTM, post-translational modification; TICA, time-sensitive independent correlation analysis; Aβ, amyloid beta; hIAPP, human islet amyloid polypeptide; EGFR, epidermal growth factor receptor.

* Corresponding author. Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain.
E-mail address: gianni.defabritiis@upf.edu (G. De Fabritiis).

http://dx.doi.org/10.1016/j.pbiomolbio.2015.03.003
0079-6107/

Extensive investigation over the last decade has made it clear that IDPs have frequent and important roles in biological processes. Disorder is found in both prokaryotes and eukaryotes, but is higher in eukaryotes, where disordered regions are found in more than 50% of proteins (Pancsa and Tompa, 2012). Disordered regions are enriched in regulatory and signaling proteins, and are less commonly found in proteins responsible for metabolism, biosynthesis or transport. They have various functional advantages, among them the ability to bind to multiple different binding partners, to form weak but highly-specific interactions, as well as being frequent targets of post-translational modification (Dunker et al., 2002; Dyson, 2011; Dyson and Wright, 2005).

Having such prevalence in key regulatory functions in the cell means that they are commonly found to play roles in various diseases (Uversky et al., 2008). They are found mutated in numerous cancers (Iakoucheva et al., 2002), unexpectedly common in cardiovascular disease (Cheng et al., 2006a), and they are common components in the fibrils of various amyloidoses like Alzheimer's disease (Lashuel et al., 2002), Parkinson's disease, and diabetes (Höppener et al., 2000). A better understanding of how and why they cause or participate in such diseases is crucial from a therapeutic perspective.

Some of the first clues to the existence of IDPs came from crystal structures with missing sections in their electron density maps (Alber et al., 1983; Lewis et al., 1996; Spolar and Record, 1994), in some cases parts critical to function (Aviles et al., 1978; Muchmore et al., 1996). This, along with data from NMR and CD experiments, led to the creation of a database for disordered regions like DisProt (Sickmeier et al., 2007) and inspired tools to try to predict disorder from sequence alone (PONDR) (Romero et al., 1997). There are now many such predictors (He et al., 2009; Linding et al., 2003), and even meta predictors (Xue et al., 2010).

Some work has already been done to understand what chemical properties make IDPs disordered, and what static or dynamic structural properties they may have. Most disordered proteins are polyampholytes, enriched in charged residues and depleted in hydrophobic residues (Uversky et al., 2000). The distribution of charge along the sequence determines how collapsed or extended the IDP is (Das and Pappu, 2013). The prevalence and importance of transient metastable structures is unclear, but contact between hydrophobic residues has already been shown as one mechanism that confers such transient structure (Meng et al., 2013). A database has been created to store structural ensembles of IDPs for future study (Varadi et al., 2013).

Despite all this progress, IDPs are still difficult to study from a biophysics point of view. Generally the methods used suffer either in limitations in their scale or time resolution. X-ray crystallography, for example, can give accurate information on atomic positions, but is limited by the fact that positions of atoms must be stable, or at least transition only slowly between a few positions. Nuclear magnetic resonance (NMR) methods can give general information about residual secondary structure or transiently formed long range contacts (Eliezer, 2009), as well as the timescale of conformational transitions. However, the information is ensemble averaged and, despite recent advances (Ban et al., 2013; Palmer III, 2014), limitations on the accessible timescale remain. SAXS is particularly well suited to study the degree of collapse in IDPs, and single-molecular FRET can single out conformations with varying degree of extension. In short, experimental methods have clear limitations in their ability to give detailed information about the states and transitions of IDPs.

The challenges faced by the methods above stress the need for new approaches. Secondary structural motifs like α-helices and β-hairpins form on the 0.1–10 μs timescale, and even the fastest folding proteins take multiple microseconds to milliseconds to fold (Lindorff-Larsen et al., 2011; Snow et al., 2002). Meaningful transitions in IDPs will likely occur on similar timescales, so any technique that is to fill this void must be able to identify transitions and metastable states formed on these timescales or longer.

Long timescale, explicit solvent molecular dynamics simulations is perhaps just the tool. Molecular dynamics simulations use a classical Newtonian representation of atoms, molecules, and the forces between them are encoded in a forcefield which contains all the chemical specificity (Karplus and McCammon, 2002; Levitt, 2001). Specialized computer hardware now exists that allow one to perform single simulations on the millisecond timescales (Ohmura et al., 2014; Shaw et al., 2007). While they are important steps forward, those tools are expensive and difficult to access. New methods pioneered by our group and others that use off-the-shelf GPU hardware and specialized analyses mean such investigations are open to a much broader research community. In the following sections, we highlight the new and important role these tools have had in investigating several disordered proteins (Table 1). While numerous kinds of simulations have been done, we focus on the state-of-the-art by looking at unbiased, all-atom, explicit solvent simulations that extend into the tens of microseconds and beyond. While many biasing or coarse-graining techniques exist to accelerate such simulations, they often require prior knowledge about a system. We direct the reader to a comprehensive review of such techniques and their limitations (Zuckerman, 2011).

2. Extensive all-atom simulations of IDPs

The first notable study of the disordered state of a protein using extensive molecular dynamics was that of bovine acyl-coenzyme A binding protein (ACBP) performed on the Anton supercomputer (Lindorff-Larsen et al., 2012). It covered 200 μs of simulation time, or two orders of magnitude longer than any previous study of an IDP. As the first study of its kind, the goal was to determine how well the simulations reproduced copious but sparse NMR data on the protein. Key to this end was the use of a state-of-the-art force field, CHARMM22*, which balances secondary structural propensity to match experimental data (Piana et al., 2011). Lindorff-Larsen et al. found that the simulations reliably reproduced several different NMR observables, including helical fraction of each residue, spectral densities and order parameters. The only major discrepancy they found was the radius of gyration, which was significantly more collapsed in the simulations.

One of the most interesting findings of the work was that they observed a conformation that formed but did not break during simulations. This highlights the difficulty with using simulations to investigate IDPs or proteins that fold on such timescales. While this will likely be an issue with simulations in the foreseeable future, it is encouraging that a simulation of that length could describe so much of the protein's behavior.

Working with single long trajectories is convenient, as analysis gets much more complex when working with many short parallel simulations. One of the biggest issues in properly characterizing the states and motions of an IDP is that conformers may be geometrically close, but distant kinetically, and vice versa. Our group struggled with this issue in one of our first works with IDPs where we studied the HIV-1 fusion peptide (Venken et al., 2013). We were unable to adequately cluster the data into meaningful states, and therefore could only make general statements about its bulk properties. This issue was resolved with the development of the TICA method (Pérez-Hernández et al., 2013; Schwantes and Pande, 2013), which allows the data to first be projected along its slowest coordinates, before being clustered into meaningful states. This makes simulations of IDPs with highly parallel methods like GPU clusters possible.

Please cite this article in press as: Stanley, N., et al., Progress in studying intrinsically disordered proteins with atomistic simulations, Progress in Biophysics and Molecular Biology (2015), http://dx.doi.org/10.1016/j.pbiomolbio.2015.03.003