Communications in Computational Chemistry
Commun. Comput. Chem., Vol. 7, No. 4, 350–360
doi: 10.4208/cicc.2025.217.02
Regular Article

Combining All-Atom Molecular Dynamics Simulation and NMR to Analyze Conformational Ensemble of Intrinsically Disordered Proteins

Xingyu Song and Wenning Wang*

Department of Chemistry, Institute of Biomedical Sciences and Multiscale Research Institute of Complex Systems, Fudan University, Shanghai 200438, China.

* Corresponding authors: wnwang@fudan.edu.cn

Received on 5 August 2025; Accepted on 2 November 2025

Abstract: Intrinsically disordered proteins (IDPs) lack stable tertiary structures and instead populate dynamic conformational ensembles, presenting unique challenges for structural characterization. In this review, we discuss the synergistic integration of all-atom molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) spectroscopy to elucidate the structural and dynamic properties of IDPs. NMR spectroscopy provides ensemble-averaged, site-specific structural and dynamic information, though its inherently sparse data limits resolution. Conversely, MD simulations yield atomically detailed trajectories but are constrained by sampling limitations and potential force field inaccuracies. Integrating both methods, using NMR data as restraints or reweighting criteria for MD simulations, improves accuracy and provides a more complete understanding of IDP behavior. Recent advancements include statistical reweighting techniques and AI-assisted methods to enhance sampling efficiency and ensemble construction. Despite progress, challenges remain in force field accuracy and seamless data integration. Future work will focus on improving force fields, developing more dynamic data integration methods, and leveraging AI for more efficient and accurate ensemble generation.

Key words: intrinsically disordered protein, molecular dynamics simulation, NMR, structure ensemble.

1. Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) defy the classical structure-function paradigm by lacking a stable tertiary structure under physiological conditions [1-5]. Instead of adopting a single, well-defined conformation, IDPs exist as dynamic ensembles of rapidly interconverting structures [6,7]. This inherent flexibility enables them to engage in diverse biological functions, including cellular signaling [8-10], transcription regulation [9,11], and molecular recognition [12], and closely associated with a range of diseases, such as neurodegenerative disorders, cancer, and cardiovascular diseases [13-15].

A central concept in the study of IDPs is that of the conformational ensemble, which provides a statistical description of all structural conformations [5,16]. Unlike folded proteins, whose structures can often be captured by single high-resolution models, IDPs require a more comprehensive depiction that accounts for their conformational heterogeneity, structural plasticity, and dynamic fluctuations [17-21]. Capturing these ensembles with sufficient accuracy is critical for understanding their biophysical behavior and functional mechanisms [22-26].

Due to their high flexibility, characterizing the conformational ensembles of IDPs remains an immense challenge [27]. Their transient and heterogeneous nature renders many conventional structural biology techniques insufficient [28]. High-resolution methods such as X-ray crystallography or cryo-electron microscopy often fail to resolve disordered regions due to their inherent flexibility [29]. In contrast, nuclear magnetic resonance (NMR) spectroscopy [30] has emerged as a powerful and complementary tool for probing IDP ensembles at atomic resolution. NMR spectroscopy is uniquely well-suited for studying IDPs, as it provides ensemble-averaged, site-specific information

on local structure and dynamics under physiological conditions. Parameters such as chemical shifts [31], residual dipolar couplings (RDCs) [32], nuclear Overhauser effects (NOEs) [33], paramagnetic relaxation enhancements (PREs), and spin relaxation rates [34] offer rich insights into both secondary structure propensities and backbone dynamics. However, NMR data are inherently sparse and indirect, requiring interpretation through structural modeling or computational support [35].

On the other hand, all-atom molecular dynamics (MD) simulations can provide atomistic trajectories of protein motions, thereby offering a detailed picture of conformational sampling and dynamics across a wide range of time scales [36-38]. However, due to the limited sampling efficiency and force field accuracy, MD simulations may deviate from physical reality or fail to capture functionally relevant states [35]. Given their respective strengths and limitations, integrating all-atom MD simulations with NMR data has become increasingly essential for constructing accurate and experimentally validated models of IDP ensembles [39,40]. By using NMR observables as restraints, validation metrics, or reweighting criteria, simulations can be refined to reflect true biophysical behaviors, while MD offers atomistic context and dynamical interpretation to otherwise averaged experimental data [41]. This synergistic approach holds great promise for unraveling the complexity of IDPs. Here, we present a review of the recent advances in the integration of MD simulations and NMR spectroscopy for characterizing the conformational ensembles of IDPs.

2. NMR characterization of IDP structure and dynamics

NMR spectroscopy yields a range of experimental observables that allow researchers to explore the structural and dynamical properties of biomolecules [42]. Crucially, the technique captures ensemble-averaged signals weighted by the population of interconverting conformational states under equilibrium [43].

2.1 Chemical shifts

Chemical shifts are fundamental parameters in NMR spectroscopy, describing the position of an NMR signal relative to a reference [44,45]. As highly sensitive reporters of molecular structure, chemical shifts have been extensively utilized in protein structure prediction, enabling accurate and efficient determination of structures based solely on chemical shift data. [45-49]. In recent years, chemical shifts have become standard tools for modelling the secondary and tertiary structures of IDPs, providing critical insights into their conformational ensembles [43,50-54]. Different types of chemical shifts exhibit distinct yet complementary dependencies on backbone dihedral angles, thereby enabling residue-specific conformational mapping of IDPs [55].

Among various chemical shift-based observables, secondary chemical shifts (SCSs)—defined as the difference between the measured chemical shift and the corresponding random coil chemical shift (RCCS)—serve as the primary atomic-scale indicators of local secondary structure propensities [56,57]:

$$SCS_i^A = \delta_i^A - RCCS_i^A \quad (1)$$

where i denotes the residue position and A indicates the atom type. Depending on the nucleus observed and the conformational space sampled (e.g., α-helical or β-sheet regions of the Ramachandran plot), SCS values can be either positive or negative. This sensitivity makes them powerful indicators of transient secondary structural elements, even within highly dynamic or disordered regions [55].

2.2 Nuclear overhauser effect (NOE)

The Nuclear Overhauser Effect (NOE) provides inter-atomic distance information, typically between hydrogen atoms within approximately 5 Å of each other [58]. NOEs arise from cross-relaxation processes between spatially close nuclear spins and are widely used as key constraints in the determination of 3D structures of biomolecules, especially proteins and nucleic acids [59-61].

Notably, in studies of IDPs, NOEs are often weak or sparse due to rapidly conformational averaging and lack of persistent tertiary contacts [62,63]. However, weak NOE signals can still provide valuable insight into local compaction, residue structure, or preferred interactions within dynamic ensembles. By scrutinizing the patterns of NOE cross-peaks, transient helices, β-strands, and other fleeting structural motifs can be identified within IDPs. Moreover, NOEs reveal long-range contacts—such as intramolecular hydrogen bonds and hydrophobic interactions—thereby mapping the sparse yet functionally relevant tertiary networks that persist in the disordered ensemble [64,65].

2.3 Paramagnetic relaxation enhancement (PRE)

PRE provides long-range distance information, by leveraging the influence of an unpaired electron—typically introduced via a paramagnetic probe—on the relaxation rates of the nuclear spins [66,67]. PRE is a useful tool for detecting long-range distance restraints, up to approximately 40 Å [68,69]. This technique is particularly valuable for investigating protein structure, dynamics, and interactions, as it can reveal sparsely populated conformational states and transient structural changes that are often inaccessible by other methods [70,71].

Because PREs are sensitive to a much wider distance range than NOEs, they are particularly valuable for detecting transient, long-range contacts, including those that exist in low-population conformational states [69,72,73]. For IDPs, PRE provides unique insight into transient tertiary interactions and compaction levels that are often invisible to other NMR observables such as chemical shifts or NOEs [74-76].

2.4 Residual dipolar couplings (RDCs)

Residue Dipolar Couplings (RDCs) are NMR observables that provide long-range internuclear orientational information [77-81]. They offer insights into the time- and ensemble-averaged orientation of bond vectors, thereby complementing high-resolution structural and dynamic studies of proteins and other biomolecules [80,82-84].RDCs are particularly valuable for refining global protein fold and domain orientation [85-88], studying conformational equilibria and dynamics [89-91], and investigating protein-protein and protein-ligand interactions [92,93]. As for IDPs, RDCs report on angular rather than distance constraints, can significantly enhance the accuracy and precision of NMR-based structure-determination, especially in cases where conventional data is sparse [94-97].