

Review Article

Synergistic Applications of MD and NMR for the Study of Biological Systems

Olivier Fiset, ^{1,2,3} Patrick Lagüe, ^{1,2} Stéphane Gagné, ^{1,2} and Sébastien Morin ⁴

¹ *Département de Biochimie, de Microbiologie et Bio-Informatique, Université Laval, Québec City, QC, Canada*

² *PROTEO and IBIS, Université Laval, Québec City, QC, Canada*

³ *Charles-Eugènes-Marchand Building, Office 4252, Université Laval, Québec City, QC, Canada G1V0A6*

⁴ *Department of Structural Biology, Biozentrum, University of Basel, Switzerland*

Correspondence should be addressed to Olivier Fiset, olivier.fiset.1@ulaval.ca

Received 13 August 2011; Accepted 19 October 2011

Academic Editor: Sergio Pantano

Copyright © 2012 Olivier Fiset et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modern biological sciences are becoming more and more multidisciplinary. At the same time, theoretical and computational approaches gain in reliability and their field of application widens. In this short paper, we discuss recent advances in the areas of solution nuclear magnetic resonance (NMR) spectroscopy and molecular dynamics (MD) simulations that were made possible by the combination of both methods, that is, through their synergistic use. We present the main NMR observables and parameters that can be computed from simulations, and how they are used in a variety of complementary applications, including dynamics studies, model-free analysis, force field validation, and structural studies.

1. Introduction

Structure and dynamics studies of proteins and other biological macromolecules involving the use of multiple techniques are rapidly becoming the norm rather than the exception. In addition to a wide array of experimental possibilities, structural studies benefit from theoretical techniques whose predictive abilities have increased tremendously due to both methodological developments and the rapid increase of available computer power. The joint use of nuclear magnetic resonance (NMR) spectroscopy [1] and molecular dynamics (MD) simulations [2] is becoming commonplace due to their high complementarity: while NMR yields highly quantitative data on dynamic processes, these data suffer from not being easily linked to unambiguously identified motions. On the other hand, MD simulations unambiguously describe atomic motions, but they are predictions impaired by force-field limitations and model approximations. Hence, combining the strengths of experimental data from NMR and simulation data from MD yields a more reliable understanding of dynamics in terms of quantities and physical description

of motions, respectively. This has an impact on our ability to study a variety of biological systems, from Alzheimer's disease-related amyloid peptides [3, 4] to the catalytic properties of antibiotic resistance enzymes [5].

Solution NMR spectroscopy allows the study of proteins in terms of both structure and dynamics. For proteins, labelling with ¹³C and ¹⁵N is typically performed and allows the observation, at atomic resolution, of most atoms, since naturally abundant ¹H is also NMR-active. Because such isotope labelling has no effect on protein structure and function, NMR can be considered a pseudo-label-free technique. Most NMR dynamics studies focus on N-H groups, which allows observation of a specific probe for most residues, with the notable exception of prolines and N-termini. A set of different experiment types are then available to probe motions on different timescales, ranging from the ps to several days (see Figure 1). However, NMR, as many other spectroscopic or bulk techniques, is limited by averaging [6] and experimental observations are thus driven by population statistics. Therefore, even though quantitative dynamics information can be derived from NMR, most of the time this

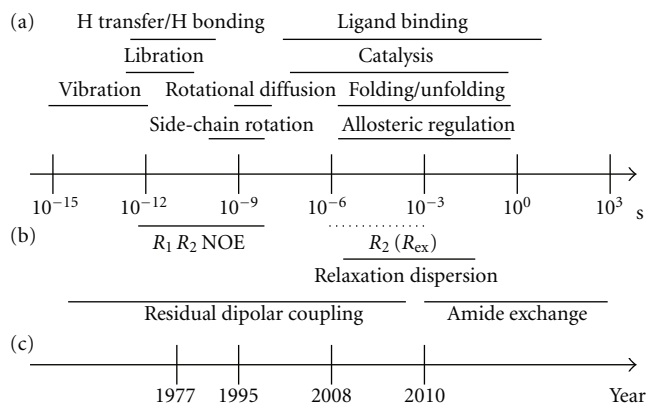


FIGURE 1: Timescales of protein motions (a) with timescales accessible to NMR experiments (b) and the approximate years these timescales became accessible to MD simulations (c).

information is limited to determining timescales (rates) of motions, rather than giving a direct physical description of these motions.

MD simulations are a structural bioinformatics technique that uses Newtonian physics to describe the dynamics of a system (such as a protein immersed in water molecules) at the atomic level. Atoms are represented by charged point masses. The force field (the expression of the potential energy as a function of atomic positions) is used to compute forces in the system and generate atomic positions and velocities along a trajectory. Modern force fields and MD implementations and techniques were recently reviewed by Guvench and MacKerell [2]. Timescales amenable to MD simulations have increased by many orders of magnitude in recent years, making MD a powerful probe of macromolecular dynamics (see Figure 1).

Case [7], in a seminal 2002 article, predicted important developments in the combined use of MD and NMR, with a focus on spin relaxation experiments applied to the study of both global rotational motions and the local dynamics of individual spins. In this short paper, different aspects of the joint use of NMR and MD are discussed. While this document is not an exhaustive listing of all the work done in the field of combined NMR and MD, it should give the reader a broad picture of how these methods can be used synergistically. Computational docking using NMR constraints, although not MD *per se*, is also discussed briefly. Our focus is on solution NMR, and therefore joint applications of MD and solid-state NMR spectroscopy (ssNMR) are not reviewed herein. (Readers interested in MD-ssNMR studies will find the recent work of Romo et al. [8] an interesting literature starting point).

In the next section, we overview methods for calculating NMR parameters and observables from MD trajectories: spin relaxation rates, order parameters, and conformational exchange. Then, applications of combined NMR and MD for dynamics studies are presented, starting with a selection of biological models that were developed using both MD and NMR. Specific applications are then treated: model-free analysis, force field validation, the special case of

backbone energetics, and long-timescale simulations. The last section summarizes the use of MD simulations in structural NMR studies, including structure determination and ligand docking. We conclude with perspectives on future usage of joint MD and NMR.

2. NMR Observables and Parameters from MD Simulations

2.1. Spin Relaxation Rates. One of the most popular approaches to the study of protein dynamics using NMR is the recording of spin relaxation rates R_1 and R_2 , as well as steady-state heteronuclear nuclear overhauser effect (NOE). R_1 is the longitudinal relaxation rate, that is, the rate with which nuclear magnetization returns to equilibrium after being perturbed. R_2 is the transverse relaxation rate, that is, the rate with which spin magnetization coherence is lost after being created. The steady-state heteronuclear NOE corresponds to the cross-relaxation between two dipolar-coupled spins [9]. These relaxation observables depend on stochastic motions on discrete timescales, which are combinations of the Larmor frequencies of the nuclear spins involved. These data can be analyzed directly, although their interpretation rarely allows the identification of the exact underlying motion.

MD can be used to calculate R_1 , R_2 , and NOE values. One approach is to compute data from spectral density values themselves computed from the partitioned correlation function reporting on the reorientation of the bond vector under study (the vector linking two bonded atoms for which motions are observed through NMR relaxation, e.g., N–H, N–C, or C–H bond vectors) [10, 11]. Alternatively, reduced spectral density values can be calculated from the experimental data and compared to spectral density values [12] extracted from MD, although the data content from such an alternative approach is reduced compared to the more direct one. Of course, these approaches, as others presented below, suffer from the limited length of trajectories, which introduces a discontinuity into the calculated correlation function [11]. This is particularly true when bond vector reorientation converges slowly, that is, the vector experiences high mobility. Although MD can be used to help in the interpretation of spin relaxation data directly, the analysis of such data is generally performed within a mathematical framework such as the model-free formalism (also known as the Lipari-Szabo formalism [13, 14], see below).

2.2. Order Parameters. In the model-free formalism (reviewed by Morin [15]), spin relaxation rates are transformed into more easily interpretable parameters such as the squared generalized order parameter (S^2), a quantitative indicator of local order, and the conformational exchange parameter (R_{ex}), which is a semiquantitative indicator of μ s-ms motions (treated in the next section). S^2 order parameters range from zero to one and are a measure of the spatial restriction of a chemical bond, in this case the N–H vector of the peptide plan. At a value of 1, there is no internal motion whatsoever in the bond vector. At 0, the bond is not constrained in any way relatively to others.

MD simulations can be used to calculate S^2 [5, 16–19]. When calculated values fit well with experimental data, MD simulations can be used with confidence for interpreting underlying motions. The classic approach to MD-derived S^2 is to compute the internal autocorrelation function for the bond reorientation. The plateau value extracted from this function at an infinite time corresponds to S^2 [17, 20] (see Figure 2). This plateau is reached when simulations converge (i.e., when the accessible conformational space is adequately sampled). Otherwise, the function is unconverged, leading to a possible S^2 overestimation. Moreover, if a single local motion dominates the trajectory, the decay to the plateau is exponential, and the area under the curve corresponds to the internal correlation time τ , the timescale of motion characterized by S^2 (see Figure 2). In a variation of this approach, the internal autocorrelation function does not use the MD data in a chronological manner, but rather randomized, with the function directly decaying to the plateau (S^2) value [5, 21]. Obviously, convergence cannot be assessed with this approach. However, it allows the combination of multiple trajectories in order to increase conformational sampling. S^2 can also be computed from spherical harmonics summation [17, 21], another technique that allows the use of multiple trajectories. Agreement between spherical harmonics and the autocorrelation function was used by Buck et al. [17] to assess convergence.

The model-free formalism was derived by assuming separation of global and local dynamics, that is, separation of timescales for the global diffusion of the protein and the local motions of the bond vector [13, 14]. For folded globular proteins, this timescale separation is generally respected. However, for unfolded proteins, where the global motions can be much faster and coupled to local movements, such separation might not be true. (It was later shown [22] that timescale separability is not an issue in most cases; a notable exception is when local motions considerably alter the overall shape of the protein.) In order to mitigate the issue of timescale separability, Prompers and Brüschweiler [16] introduced a method called isotropic reorientational eigenmode dynamics (iRED), which relies on the use of MD simulations for the analysis of experimental NMR spin relaxation data. In addition to being completely unaffected by timescale separability (and thus well suited for the study of unfolded proteins), this method can also evaluate if global and internal motions are statistically separable. The method uses principal component analysis (PCA) [23] evaluated from MD simulations in order to extract reorientational eigenmodes and amplitudes. The distribution of eigenvalues allows the assessment of timescale separation. Finally, iRED parameters are fitted to experimental data, allowing the extraction of dynamics information, such as S^2 values. The method was applied to the iron responsive element RNA by Showalter et al. [24], who could demonstrate that global and local motions are not statistically separable, but rather correlated for this small macromolecule. Maragakis et al. [19] also proposed a solution to the separability problem, where the full autocorrelation function (as opposed to the local one) is used, thereby recoupling global and local motions to simulate the effect of global tumbling on NMR

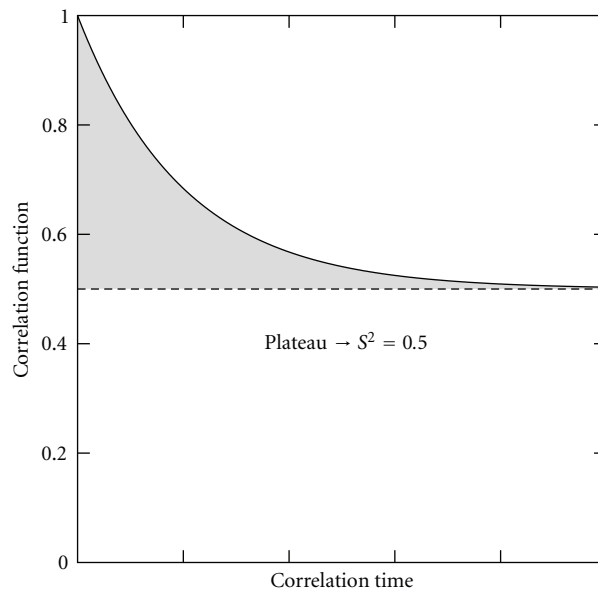


FIGURE 2: Internal autocorrelation function for an imaginary bond vector where $S^2 = 0.5$, the plateau value as the function converges towards infinity. The area between the curve and S^2 is the local correlation time (τ).

relaxation. Finally, Johnson et al. [25] created different helical ensembles by modification of ϕ/ψ angles in order to compare different approaches to the extraction of order parameters for multiple different backbone bond vectors. Their approach highlighted the fact that multiple analytical methods applied to the same system can yield a complementary and self-consistent picture of reorientational motions. In other words, extracting S^2 using different approaches allows inferences on the underlying motions in action, whereas a single S^2 value (from a single approach) does not provide information on the physical nature of this motion.

Whereas order parameters extracted from spin relaxation data (using either the model-free formalism or the iRED approach) report on the ps-ns timescale, residual dipolar couplings (RDC, couplings observable for weakly aligned samples) can be used to extract a broader timescale S^2 ranging from ps to ms [26–28] (see Figure 1). The difference between spin relaxation and RDC S^2 can highlight the presence of slower μs -ms movements. Motions on this physiologically relevant timescale can improve the understanding of protein function as they might be coupled to activity [29]. As for the analysis of model-free order parameters, the analysis of RDC S^2 data also can benefit greatly from combination with MD. Indeed, Salmon et al. [30] very recently reviewed this particular application, highlighting the fact that amplitudes of motions, as well as their distribution, can reliably be extracted from the two techniques (in this case, with accelerated MD simulations, discussed further below), allowing cross-validation of both techniques and increasing the confidence with which these data can be used for the understanding of protein motions.

2.3. *Conformational Exchange (R_{ex})*. μ s-ms motions, often related to protein functions such as enzyme catalysis [29], can be indirectly studied using a combination of order parameters extracted from spin relaxation and RDC (see above). A more direct approach is the extraction of a conformational exchange term (R_{ex} , also called chemical exchange) either from model-free analysis of relaxation data or from relaxation dispersion experiments. R_{ex} arise from μ s-ms motions because these affect the magnitude of the effective transverse relaxation rate R_2 . R_{ex} values depend on timescales of motion, chemical shift differences between different conformers, and population weight of these conformers. R_{ex} sums with the pure R_2 (reporting on ps-ns motions) to give the observed R_2 [31]. However, interpreting R_{ex} in terms of physical motions is not trivial (with some exceptions [32–35]). MD simulations could be useful in interpreting experimental R_{ex} parameters. Unfortunately, the μ s-ms timescale is currently barely accessible to all-atom MD. This might soon change as simulations reaching the ms timescale are now being produced [36] (see Figure 1), although on special purpose hardware that is not commonly available. Once this timescale is routinely simulated, it will be possible to compare R_{ex} values with observed slow motions qualitatively, and possibly quantitatively by correlating R_{ex} values with timescales of motion for conformers with distinct chemical shifts. Such comparisons will be facilitated by the high accuracy with which chemical shifts can now be predicted from 3D structure (reviewed by Wishart [37]).

Alternatively, accelerated molecular dynamics (AMD) techniques can be used to probe the μ s-ms timescale. AMD are a variety of techniques designed to efficiently sample low-energy conformations. This implies escaping energy minima, often by artificially reducing the energy barrier between low-energy conformations. AMD techniques were reviewed by Elber [38]. Markwick et al. [21] showed that AMD can be used together with NMR spectroscopy to study multiple timescale motions in proteins. They used AMD simulations of GB3 protein to quickly sample motions on the ms timescale. Classic MD simulations were then performed using AMD conformations as starting points, providing insights into dynamics at different timescales (ms versus ns). NMR relaxation and RDC data available for that protein were in agreement with their findings.

3. Dynamics Studies

There are many examples in the literature of practical applications of NMR and MD for the study of biological systems. MD simulations are often used to facilitate the interpretation of NMR dynamic experiments, and to find new leads that can then be pursued experimentally; MD can also be used to perform experiments that are inaccessible to NMR and to focus on a single molecule rather than an ensemble average. In turn, NMR validates simulations of specific systems and thus their predictive power. Examples are discussed below.

The 36 residue villin headpiece helical subdomain (HP36) was studied in its unfolded state by Wickstrom et al. [39]. Local structure in the unfolded state is thought to be

important to protein folding. However, it is difficult to measure experimentally because those structures convert rapidly to the folded state. An approach to study local structure is the use of peptide fragments. HP36 is made of three α helices. Corresponding fragments (HP-1, 2, and 3) were studied by NMR, where a small helical propensity was detected in HP-1 and 3. However, the presence of multiple conformations at equilibrium leads to ambiguities in data interpretation. To palliate limitations of the experimental methods, the fragments were simulated using replica-exchange molecular dynamics (REMD), an accelerated sampling technique [40]. REMD facilitates overcoming energy barriers and thus makes it possible to analyze folding, an event that would normally happen on a timescale much longer than that accessible to standard simulations (see Figure 1). Cluster analysis of the fragment conformations shows that there is a local stabilized structure in HP-1 identical to the helix observed in complete HP36. Fragments HP-2 and HP-3, on the other hand, show little helicity. Cooperative folding in HP36 could, therefore, start by local helix formation in the HP-1 region. These simulations are validated by the agreement between experimental and simulated J-couplings.

Using implicit solvent, Kent et al. [3] developed a model of the aggregation of β -amyloid fragments ($A\beta$). Since $A\beta$ aggregates on μ s or longer timescales, simulations of such atomistic models are expensive. Using NMR order parameters for peptide $A\beta(10-35)$ and a 10 ns explicit solvent simulation of the same fragment as reference data, they benchmarked the accuracy of force fields CHARMM, OPLS-AA, and Amber combined with their Shen-Freed implicit solvent model. Order parameters varied considerably with force field, but CHARMM22 displayed good agreement with experiment. After a 30 ns equilibration time, their $A\beta$ model exhibits high flexibility, consistent with experiment, and maintains the strand-loop-strand motif that promotes aggregation through a solvent-exposed hydrophobic patch. Further work on $A\beta$ peptides using both MD and NMR was performed by Fawzi et al. [4]. They used new high-field experiments on the $A\beta(21-30)$ fragment to validate simulations using the Amber ff99SB force field with the TIP4P-Ew water model. They found that ^{13}C relaxation parameters are in good agreement with predicted values. Their explicit solvent model describes accurately the structure and dynamics of the complete $A\beta(21-30)$ peptide.

Ferner et al. [41] studied the dynamics of two YNMG RNA tetraloops at different temperatures using both NMR and MD simulations. This represents a particular challenge for both techniques. With NMR, careful choice of temperature-dependent parameters (such as bond length and CSA) is required for model-free analysis. MD force fields are parametrized to reproduce room temperature physical properties, so the actual effect of simulation temperature on protein dynamics is hard to predict. The thermal stability and biological function of both loops is known to be temperature dependent. Even though agreement between experimental and MD-derived S^2 decreases as the temperature is increased above room temperature, the authors are still able to observe, both in experiment and simulations, the same loss of stacking interactions between first and third bases that

leads to RNA melting. RNA dynamics were also studied by a joint MD-NMR approach by Frank et al. [42]. They generated ensembles of RNA conformations for two variants of the transactivation response element (TAR). They did so by selecting snapshots from MD trajectories according to the fit between simulated and experimental RDCs. The dynamic ensembles obtained showed that ligand binding to TAR involves an adaptive recognition mechanism and important conformational changes in RNA.

Fisette et al. [5] used NMR relaxation and order parameters to validate β -lactamase TEM-1 simulations. In turn, simulations and PCA were used to interpret relaxation data in terms of physical motions. Multiple simulations were shown to be necessary to sample the conformational space of flexible loops. Simulations show their limits in the catalytically-relevant Ω loop, where motions on slow timescales (μ s-ms) that promote substrate gating cause discrepancies between NMR and MD S^2 . In more recent work (Fisette et al., in preparation), effects of substrate binding on TEM-1 and PSE-4 class A β -lactamases were studied by MD simulations, overcoming a limitation of NMR experiments: rapid enzyme turnover makes relaxation experiments in presence of saturating substrate levels impossible; the only experimental workaround is the use of covalent inhibitors or variant enzymes, which would significantly affect dynamics around the active site.

Interested in the dynamics of a small flexible RNA fragment, Musselman et al. [43] used a domain-elongation approach to perform different NMR spin relaxation experiments, and compared their results with MD-predicted relaxation (R_1 , R_2 , and NOE), as well as model-free (S^2 and τ) parameters. The domain-elongation approach developed by Zhang et al. [44] consisted in adding several Watson-Crick base pairs on one side of the RNA under study using unlabelled (i.e., NMR-invisible) nucleotides to slow down global diffusion and, thus, both decouple local and global motions and allow a higher sensitivity to fast dynamics. For Musselman et al. [43], this approach was useful in allowing a defined reference frame for analysis of MD, an otherwise difficult task with such a flexible entity where local and global motions are coupled. The RNA displayed a complex mixture of local and global dynamics on multiple timescales. Whereas the fast motions were detected by both spin relaxation and MD, slower motions were indicated by MD as well as RDC data.

3.1. MD-Guided Model-Free Analysis. When applying the model-free formalism, a model selection step is necessary to decide on the inclusion of different parameters and timescales, the basic idea being to identify the most statistically-relevant model and parameter set yielding a good fit of the experimental data. Model selection protocols and parameter accuracy of the model-free formalism were studied by Chen et al. [11] using MD simulations. They generated a 10 ns trajectory of dihydrofolate reductase ternary complex and computed relaxation parameters (R_1 , R_2 , and NOE) from it. Afterwards, they used model-free analysis and compared different model selection approaches to extract dynamic parameters from the artificial relaxation data.

These parameters were then compared to those computed directly from the trajectory using autocorrelation functions. Simulations thus served as a reference where the motions are perfectly known, while the model-free analysis had to reconstruct the spectral density from R_1 , R_2 , and NOE, measurements that provide only partial information on the dynamics. Their results showed that, as proposed by d’Auvergne and Gooley [45], the use of not only Bayesian Information Criteria (BIC) [46] but also Akaike Information Criteria (AIC) [47] in the case where multiple magnetic field data are recorded are superior compared to the then generally used step-up hypothesis approach involving F -tests [48]. Indeed, these information theoretical criteria provide a better balance between bias and variance (less under- and over-fitting) and lead to more reliable model selection, while also being less computer intensive.

MD simulations can also play a direct role in model selection, as was recently shown by Fisette et al. [5]. In this approach, model selection proceeds as usual with statistical tests for the comparison of experimental and back-calculated values, but is helped for cases where the tests yield a similar statistical score for different models (i.e., when different models reproduce equally the experimental data). In these ambiguous situations, projections of the orientation of the bond vector in the course of MD simulations are used to evaluate the motion type and decide on the best model. Even though this approach should, in principle, allow the correct identification of local motions, it suffers from two problems with regards to the R_{ex} parameter. The first is obviously the timescale probed by the trajectory. The second is the case where the R_{ex} contribution arises not from motions in the bond vector itself, but rather from movements in nearby atoms that affect the electronic environment of the bond vector on the μ s-ms timescale; such motions are undetected in the projection of a given bond vector.

3.2. Force Field Validation. If we are to trust the predictive capabilities of a given force field, that force field should reproduce experimental observations within its range of application. The classical, additive force fields used for MD simulations of macromolecules are developed using quantum mechanics (QM) properties (geometry, vibrational modes, solvation energies, etc.) and experimental data. Unfortunately, the direct fitting of a force field to proteins or even small peptides is impossible since such systems are not tractable by QM calculations at the required theory level. Instead, final parameter sets are used for protein simulations that are then compared to experimental results. Hen egg white lysozyme [17, 18, 49–52] and ubiquitin [19, 50–53] are often used as model systems since they are small and very well characterized experimentally. NMR observables used for force field validation include root-mean-square deviation (RMSD) from NMR structural ensembles [49], NOE (measurement of inter-proton distances) [49, 54], scalar spin-coupling constants [52, 55, 56], and chemical shifts [57, 58]. Direct comparisons of spin relaxation parameters [53] are a lot less frequent. Instead, relaxation-derived quantities such as S^2 [5, 17–19, 49, 50, 53, 59–61] are used. RDC-derived

order parameters [52, 60, 62] are increasingly used for long-timescale simulations.

A detailed validation of a GROMOS force field parameter set (45A3) by Soares et al. [49] used a variety of NMR data and two 3.5 ns simulations of lysozyme. They used a crystallographic structure as the simulations starting point. Interestingly, trajectory RMSD shows that the trajectories stay closer to the X-ray structure than to the lowest-energy NMR conformer. This is probably indicative of trajectory dependence on initial conditions, a topic that was further studied by combined NMR-MD studies, and which will be discussed later on. Soares et al. [49] also compared predicted NOE with experimental values. Their results suggested the new GROMOS parameter set violated experimental NOE constraints slightly more than the previous set (43A1). However, studies by Zagrovic and van Gunsteren [54] later showed that comparing NMR and MD-predicted NOE may not be a good indicator of simulations quality due to the nonlinear effect of r^{-3} or r^{-6} averaging [63] and focus on experimentally observed NOE that produces artificially good agreement between the two techniques. Predicted spin-coupling constants were also used by Soares et al. [49] to assess force field quality. Both three-bond ${}^3J_{HN\alpha}$ - and ${}^3J_{\alpha\beta}$ -coupling constants converged to experimental results for most residues. For some, however, trajectory length was not sufficient to achieve a meaningful comparison. Lack of conformational space sampling in MD simulations to reproduce NMR observables is also an important issue, which will be presented below.

3.3. Backbone Energetics. An important limitation of MD simulations is that, as longer timescales become accessible, force field quality, integrator stability, and other problems that were not apparent before can be revealed and negatively impact simulations quality. N–H-squared order parameters S^2 were instrumental in validating the ϕ and ψ dihedral angle corrections necessary for common protein force fields when simulations began probing the many nanoseconds timescale. Before that correction, classical force fields would overestimate the flexibility of secondary structures. In some cases [64], short helical peptides would fold as π helices rather than the experimentally proven α helices. Cross-map term corrections (CMAP) were developed by MacKerell Jr et al. [65] for the CHARMM force field to account for secondary structure stability while keeping other parameters unchanged. CMAP is a detailed grid of the dihedral potential of ϕ and ψ angles that was integrated to the CHARMM22 force field as a supplemental term. A 25 ns lysozyme trajectory acquired by Buck et al. [17] compared predicted N–H S^2 to experimental ones. Results showed that CHARMM22/CMAP predicted S^2 that were consistent with experiment, while in absence of CMAP secondary structures were much too flexible.

Similar corrections were performed for the AMBER force field: in the ff99SB parameter set, ϕ and ψ dihedral angles were reparametrized [50] to achieve a better balance of secondary structures. Validation included a comparison of experimental and predicted S^2 for lysozyme and ubiquitin, showing much better agreement for the new parameters.

Showalter and Brüschweiler [53] also used ubiquitin simulations to validate AMBER ff99SB, but compared experimental and back-calculated relaxation data in addition to relaxation-derived order parameters. Showalter et al. [59] also showed that the modified backbone energetics in ff99SB improved the description of amino acid methyl side-chain rotation, suggesting a direct relation between main- and side-chain dynamics.

Even with this correction, however, modern force fields for proteins are still scrutinized through comparison with NMR results, revealing limitations. For instance, discrepancies in predicted and experimental J-couplings led Best et al. [55] to suggest that current force fields might be too helical as they overpopulate the α conformation of polyalanine peptides in water. Another study of short peptides by Wickstrom et al. [56] used NMR scalar coupling constants as reference data and also showed shortcomings in Amber ff99SB for Ala3 and Ala5 peptides and the TIP3P and TIP4P-Ew water models. Another NMR-MD comparison by Trbovic et al. [60] highlighted deficiencies in the OPLS-AA and AMBER (ff99SB and ff03 parameter sets) force fields. Using over 50 short (2.4 ns) simulations of the B3 immunoglobulin-binding domain of streptococcal protein G (GB3), they compared MD-predicted S^2 to experimental values determined by both spin relaxation and RDC experiments. Their results indicate that the three force fields overestimate the flexibility of amide N–H vectors at the borders of secondary structure elements and in loops. They suggest that an imbalance between hydrogen bonding and other terms may be the cause of the problem. Even though these studies concluded the impact on the accuracy of typical protein simulations is small, NMR-MD studies of small peptides may lead to force field improvements in the future.

3.4. Diffusion and Long-Timescale Simulations. Molecular dynamics rely on the ergodic hypothesis for simulating physical parameters; adequate conformational space sampling on the timescales of the biological processes being studied is required. This is rarely possible for large macromolecules, as is clearly shown by trajectory dependence on starting coordinates. For lysozyme, Koller et al. [18] showed that starting structure had a significant impact on MD-derived order parameters. Since those are increasingly used to assess force field quality, caution must be taken to acquire trajectories at least an order of magnitude longer than the solute global tumbling time. Genheden et al. [61] also studied the effects of starting structures, trajectory length, and S^2 calculation method on MD-derived S^2 . Using the carbohydrate-binding domain of galectin 3 in the free and lactose-bound states, they compared predicted order parameters with experimental data from NMR spin relaxation. They conclude that the use of multiple trajectories (ten) of a length of 125% of global tumbling (10 ns in their case) yields the best results. In both studies, the improvement obtained from longer or multiple trajectories was significant only for a small number of residues that are either very flexible (with $S^2 < 0.7$) or involved in slow dynamic events. Similar results

were observed by Fiset et al. [5] in a joint MD-NMR study of TEM-1 β -lactamase.

With MD trajectories now routinely reaching 100 ns, the global tumbling of small proteins is becoming an object of study in simulations. Global tumbling is easily measurable using solution NMR spectroscopy. Wong and Case [51] made an extensive study of three water models (SPC/E, TIP4P-Ew and TIP3P) and four proteins (ubiquitin, binase, lysozyme, GB3). Extracting the diffusion tensor for these solutes required simulations of 20 to 100 times longer than the rotational tumbling. Their study highlights how current water models make protein diffusion too fast in simulations. The diffusion constant of the popular TIP3P model is known to be about twice that of water at room temperature, but even for SPC/E and TIP4P-Ew models, whose diffusion constants are nearer experimental values, the rotational diffusion constants of all studied proteins were too large when compared to NMR results. In the first published work involving a μ s trajectory, Maragakis et al. [19] also observed a diffusion constant about twice the experimental value for ubiquitin using SPC water molecules. Now that even longer timescales are becoming accessible, theoretical approaches need to be improved once again, this time with optimized water models. However, major changes in water models are likely to require modifications to existing forcefields, since these potentials are interdependent. Water models and their limitations were reviewed by Guillot [66].

Finally, simulation stability on the μ s timescale has received attention in recent work by Lange et al. [62]. They used NMR RDC experiments, which are sensitive to dynamics from ps to ms, and compared them with μ s simulations of ubiquitin and GB3 for a variety of force fields. Interestingly, they noted that past 25 ns, increasing trajectory length improves RDC fit only marginally. In some cases, longer trajectories yield worse synthetic RDCs, suggesting that the accumulation of force-field imprecision outweighs the increased conformational space sampling and the protein behavior starts to deviate from its native state. PCA [23] was used on all MD ensembles to find correlations between RDC mismatches and structural changes. They showed that after several hundred ns, high free energy states are overrepresented and transitions to nonnative states happen at a higher frequency. They also demonstrated that the use of particle mesh Ewald (PME) for the treatment of electrostatics is necessary to obtain an accurate RDC fit. Recently, two new sets of parameters for the Amber force field were developed using NMR observables, with the explicit goal of improving the quality of long-timescale simulations. The first one, named ff99SB-ILDN, was developed by Lindorff-Larsen et al. [52] and is limited to reparametrization of χ_1 dihedral angles. They used simulations of small model peptides to identify the residue types with rotamer populations that differed most from expectations based on statistical analysis of the Protein Data Bank (PDB). Full scans of the potential energy surface (PES) of the χ_1 angles of these residues were then performed. A new torsion term in the Amber force field was introduced and replaced the previous one, and parameters for that term were fitted using the QM results. Afterwards, validation was performed with globular proteins (lysozyme,

BPTI, ubiquitin, and GB3) by comparing calculated 3 J scalar couplings and side-chain RDCs to experimental data from the literature, showing noticeable improvements over ff99SB. The second set of parameters is the work of Li and Brüschweiler [57, 58] and uses NMR chemical shifts from the BioMagResBank (BMRB) [67] as a reference for direct force field parametrization. This is a novel approach, as NMR observables are usually used for validation only, because systematic exploration of the parameter space is too expensive when simulating entire proteins. Here, the authors used an initial trajectory that was reweighted for each new parameter set using Boltzmann's relation. Parameter fitting was done by comparing theoretical and experimental chemical shifts of C, C α , and C β atoms of each residue and adjusting the potential of backbone dihedral angles ϕ and ψ . Very recently, Li and Brüschweiler [68] extended their method to use RDC in a similar way, either alone or in combination with chemical shifts. The resulting parameter sets perform better for reproducing NMR data and exhibit lower RMSD between trajectories and X-ray crystallographic structures.

As the timescales amenable to MD simulations continue to increase, and ms simulations become possible [36], NMR will most certainly continue to serve as a powerful experimental reference for force field validation and development.

4. Structural Studies

Protein structure determination with NMR data, since its first uses, has been helped by computational approaches, such as distance geometry optimization [69]. Simulated annealing approaches have been used for the calculation of structures based on NMR constraints (NOEs, paramagnetic relaxation enhancements (PREs), chemical shifts, coupling constants, RDCs, etc.) [70]. The approach consists of incorporating the experimental constraints and weighing them with a given force field in simulations performed at high effective temperature (to reduce energy barriers and allow the structure to relax and minimize its energy and the violation of experimental constraints). The temperature is then lowered slowly until a converged structure satisfying both experiment and physical constraints is obtained. In the end, a family (group) of low energy conformers (generally 10–40) is used to represent the structure. This approach and more modern implementations have been available in programs such as CYANA (DYANA) [71–74], XPLOR-NIH [75, 76], CNS [77, 78], and ARIA [79, 80]. Recently, MD has, however, been used in an increasing number of approaches different from strict NMR-constraints-based structure calculation, with some examples given below.

Three-dimensional protein structure prediction has long been one of the ultimate goals of theorists. A very successful approach has been with the program ROSETTA [81] from the Baker lab which implements *ab initio* approaches. Whereas first implementations were solely based on theory (not using experimental data), newer implementations use raw NMR data such as chemical shifts (CS-ROSETTA)

[82, 83], a combination of backbone chemical shifts and unassigned NOE data (CS-DP-ROSETTA) [84], or a combination of backbone chemical shifts, amide proton distances (from NOE data) and RDC data (CS-RDC-ROSETTA) [85]. This combination of *ab initio* approaches with easily obtained NMR data (mostly excluding the tedious assignment of side-chain resonances) allows rapid and accurate structure predictions to be made, a good compromise to full structure determination.

NMR data has also been used in conjunction with already available structures for helping the docking of ligands to potential binding sites [86]. The HADDOCK (High Ambiguity Driven protein-protein DOCKing) approach has been particularly successful in this [87, 88], first by using solely chemical shift perturbation data, but later including RDC data [89], as well as diffusion anisotropy data extracted from spin relaxation experiments [90] or pseudocontact shifts [91]. Morelli et al. [92] proposed a combined MD-NMR approach for the study of protein complexes. Named restrained soft-docking, the method uses chemical shift perturbation as a guide for a specific docking algorithm (BiGGER). Flexibility of amino acid side chains at the protein-protein interface are taken into account, but global fluctuations are disallowed, making the method most appropriate for small and medium complexes of globular proteins. Results showed excellent agreement (RMSD) with experimentally elucidated complexes for a selection of complexes: EIN/HPr, Barnase/Barnstar, Tom20/Presequence, Cyt *c*/Ccp. These complexes have interfaces that range from about 1000 to 2200 Å². The HSQC NMR experiment generally used to determine chemical shift perturbation typically requires about an hour. Combined with the performance of the docking algorithm, this makes the technique a powerful tool for structural genomics projects and the search for protein-protein complexes that can facilitate the identification of lead compounds for drug design. Developments of combined NMR and docking methods were reviewed by Morelli and Rigby [93].

The description of unfolded proteins is a particularly challenging area where, again, NMR and MD are very complementary. In this context, NMR observables are used for the validation of ensembles of conformers which are aimed at representing the situation of unfolded proteins in solution. Chemical shifts [94], scalar couplings [95], RDCs [96, 97], and PRE effects [98, 99] are different NMR observables which can be effectively used in such approaches. Different techniques exist such as the flexible-meccano algorithm by Bernadó et al. [100], in which random conformational sampling based on amino acid propensity and side-chain volume is used for the construction of typically 100,000 conformers. Another random sampling approach is that of Jha et al. [101] using a self-avoiding statistical coil model based on backbone conformational frequencies found for a subset of protein structures deposited in the Protein Data Bank (PDB). The ENSEMBLE approach [102, 103] uses Monte Carlo approaches and MD simulations in order to create relatively small ensembles of conformers for which the populations weights are then optimized using experimental

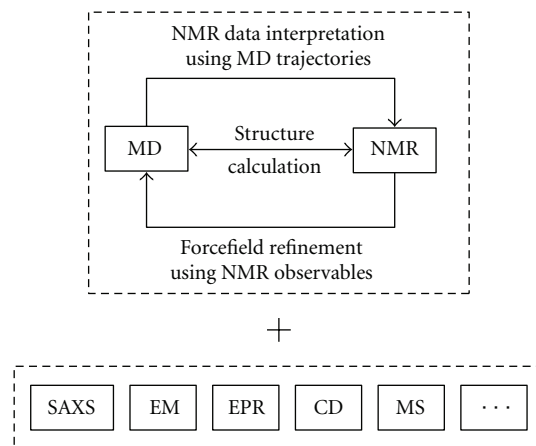


FIGURE 3: Summary of the relationship between solution NMR spectroscopy and MD simulations, as presented in this paper, along with a listing of other complementary experimental techniques.

data. The maximum occurrence approach (termed MaxOcc) [104, 105] also uses weights (in this case defined as the maximum fraction of time a conformer can exist such that experiment data is well reproduced). The originality of this approach lies in the fact that only one conformer is evaluated within the ensemble in order to avoid overfitting of the limited experimental data. Such approaches that model the distribution of conformers are very useful for the study of unfolded proteins. In particular, the combination with NMR is very effective at characterizing the persistence of secondary structures as well as hydrophobic clusters in unfolded states [106–109].

5. Conclusions

The variety of applications covered in this paper, the importance of NMR developments such as RDC, and the increase of the timescales tractable by MD simulations indicate that the joint usage of MD and NMR has a bright future. NMR will be indispensable in the development and refinement of force fields and water models able to accurately simulate both global and local protein motions on the μ s-ms timescales. MD is already an integral part of NMR structure determination, and simulations will become commonplace as a complement to NMR dynamics from relaxation or RDC to facilitate data interpretation (see Figure 3). The overall availability of both experimental and simulated data in databases such as the BMRB [67] for NMR and the Dymeomics repository [110, 111] for MD will facilitate further comparisons and enhance synergy. Moreover, approaches aimed at helping experimentalists use further computational approaches (see, for example, the worldwide e-Infrastructure for NMR and structural biology: WeNMR—<http://www.wenmr.eu/>) will also be beneficial.

Obviously, not only NMR and MD can be combined synergetically. Many other techniques such as small angle X-ray scattering (SAXS), electron microscopy (EM), electron paramagnetic relaxation (EPR), circular dichroism (CD), and mass spectrometry (MS) can be used together (see

Figure 3) to enhance their total information content as recently reviewed by Cowieson et al. [112] and exemplified with the recent study of RANTES/CCL5 oligomers by Wang et al. [113].

Acknowledgments

The authors are grateful to Simon Bernèche for constructive comments. The EMBO is acknowledged for a postdoctoral fellowship to S. Morin.

References

- [1] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, 2nd edition, 2007.
- [2] O. Guvench and A. D. MacKerell, "Comparison of protein force fields for molecular dynamics simulations," *Methods in Molecular Biology*, vol. 443, pp. 63–88, 2008.
- [3] A. Kent, A. K. Jha, J. E. Fitzgerald, and K. F. Freed, "Benchmarking implicit solvent folding simulations of the amyloid β (10–35) fragment," *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6175–6186, 2008.
- [4] N. L. Fawzi, A. H. Phillips, J. Z. Ruscio, M. Doucleff, D. E. Wemmer, and T. Head-Gordon, "Structure and dynamics of the A β _{21–30} peptide from the interplay of NMR experiments and molecular simulations," *Journal of the American Chemical Society*, vol. 130, no. 19, pp. 6145–6158, 2008.
- [5] O. Fiset, S. Morin, P.-Y. Savard, P. Lagüe, and S. M. Gagné, "TEM-1 backbone dynamics—insights from combined molecular dynamics and nuclear magnetic resonance," *Biophysical Journal*, vol. 98, no. 4, pp. 637–645, 2010.
- [6] D. Kruschel and B. Zagrovic, "Conformational averaging in structural biology: issues, challenges and computational solutions," *Molecular BioSystems*, vol. 5, no. 12, pp. 1606–1616, 2009.
- [7] D. A. Case, "Molecular dynamics and NMR spin relaxation in proteins," *Accounts of Chemical Research*, vol. 35, no. 6, pp. 325–331, 2002.
- [8] T. D. Romo, L. A. Bradney, D. V. Greathouse, and A. Grossfield, "Membrane binding of an acyl-lactoferricin B antimicrobial peptide from solid-state NMR experiments and molecular dynamics simulations," *Biochimica et Biophysica Acta*, vol. 1808, no. 8, pp. 2019–2030, 2011.
- [9] A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, UK, 1961.
- [10] R. Brüschweiler, B. Roux, M. Blackledge, C. Griesinger, M. Karplus, and R. R. Ernst, "Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution," *Journal of the American Chemical Society*, vol. 114, no. 7, pp. 2289–2302, 1992.
- [11] J. Chen, C. L. Brooks, and P. E. Wright, "Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation," *Journal of Biomolecular NMR*, vol. 29, no. 3, pp. 243–257, 2004.
- [12] J.-F. Lefevre, K. T. Dayie, J. W. Peng, and G. Wagner, "Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions," *Biochemistry*, vol. 35, no. 8, pp. 2674–2686, 1996.
- [13] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4546–4559, 1982.
- [14] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results," *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4559–4570, 1982.
- [15] S. Morin, "A practical guide to protein dynamics from 15N spin relaxation in solution," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 59, no. 3, pp. 245–262, 2011.
- [16] J. J. Prompers and R. Brüschweiler, "General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation," *Journal of the American Chemical Society*, vol. 124, no. 16, pp. 4522–4534, 2002.
- [17] M. Buck, S. Bouguet-Bonnet, R. W. Pastor, and A. D. MacKerell, "Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme," *Biophysical Journal*, vol. 90, no. 4, pp. L36–L38, 2006.
- [18] A. N. Koller, H. Schwalbe, and H. Gohlke, "Starting structure dependence of NMR order parameters derived from MD simulations: implications for judging force-field quality," *Biophysical Journal*, vol. 95, no. 1, pp. L04–L06, 2008.
- [19] P. Maragakis, K. Lindorff-Larsen, M. P. Eastwood et al., "Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins," *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6155–6158, 2008.
- [20] M. Buck and M. Karplus, "Internal and overall peptide group motion in proteins: molecular dynamics simulations for lysozyme compared with results from X-ray and NMR spectroscopy," *Journal of the American Chemical Society*, vol. 121, no. 41, pp. 9645–9658, 1999.
- [21] P. R. L. Markwick, G. Bouvignies, and M. Blackledge, "Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy," *Journal of the American Chemical Society*, vol. 129, no. 15, pp. 4724–4730, 2007.
- [22] B. Halle, "The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids," *Journal of Chemical Physics*, vol. 131, no. 22, Article ID 224507, 2009.
- [23] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, "Essential dynamics of proteins," *Proteins*, vol. 17, no. 4, pp. 412–425, 1993.
- [24] S. A. Showalter, N. A. Baker, C. Tang, and K. B. Hall, "Iron responsive element RNA flexibility described by NMR and isotropic reorientational eigenmode dynamics," *Journal of Biomolecular NMR*, vol. 32, no. 3, pp. 179–193, 2005.
- [25] E. Johnson, S. A. Showalter, and R. Brüschweiler, "A multifaceted approach to the interpretation of NMR order parameters: a case study of a dynamic α -helix," *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6203–6210, 2008.
- [26] J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, and R. Brüschweiler, "Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins," *Journal of the American Chemical Society*, vol. 123, no. 25, pp. 6098–6107, 2001.
- [27] J. R. Tolman, "A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings

- using several oriented media in biomolecular NMR spectroscopy,” *Journal of the American Chemical Society*, vol. 124, no. 40, pp. 12020–12030, 2002.
- [28] N. A. Lakomek, T. Carlomagno, S. Becker, C. Griesinger, and J. Meiler, “A thorough dynamic interpretation of residual dipolar couplings in ubiquitin,” *Journal of Biomolecular NMR*, vol. 34, no. 2, pp. 101–115, 2006.
- [29] E. Z. Eisenmesser, O. Millet, W. Labeikovsky et al., “Intrinsic dynamics of an enzyme underlies catalysis,” *Nature*, vol. 438, no. 7064, pp. 117–121, 2005.
- [30] L. Salmon, G. Bouvignies, P. Markwick, and M. Blackledge, “Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales,” *Biochemistry*, vol. 50, no. 14, pp. 2735–2747, 2011.
- [31] O. Millet, J. P. Loria, C. D. Kroenke, M. Pons, and A. G. Palmer, “The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale,” *Journal of the American Chemical Society*, vol. 122, no. 12, pp. 2867–2877, 2000.
- [32] D. McElheny, J. R. Schnell, J. C. Lansing, H. J. Dyson, and P. E. Wright, “Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 14, pp. 5032–5037, 2005.
- [33] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright, “The dynamic energy landscape of dihydrofolate reductase catalysis,” *Science*, vol. 313, no. 5793, pp. 1638–1642, 2006.
- [34] A. J. Baldwin and L. E. Kay, “NMR spectroscopy brings invisible protein states into focus,” *Nature Chemical Biology*, vol. 5, no. 11, pp. 808–814, 2009.
- [35] A. K. Mittermaier and L. E. Kay, “Observing biological dynamics at atomic resolution using NMR,” *Trends in Biochemical Sciences*, vol. 34, no. 12, pp. 601–611, 2009.
- [36] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen et al., “Atomic-level characterization of the structural dynamics of proteins,” *Science*, vol. 330, no. 6002, pp. 341–346, 2010.
- [37] D. S. Wishart, “Interpreting protein chemical shift data,” *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 58, no. 1-2, pp. 62–87, 2011.
- [38] R. Elber, “Long-timescale simulation methods,” *Current Opinion in Structural Biology*, vol. 15, no. 2, pp. 151–156, 2005.
- [39] L. Wickstrom, A. Okur, K. Song, V. Hornak, D. P. Raleigh, and C. L. Simmerling, “The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure,” *Journal of Molecular Biology*, vol. 360, no. 5, pp. 1094–1107, 2006.
- [40] Y. Sugita and Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding,” *Chemical Physics Letters*, vol. 314, no. 1-2, pp. 141–151, 1999.
- [41] J. Ferner, A. Villa, E. Duchardt et al., “NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops,” *Nucleic Acids Research*, vol. 36, no. 6, pp. 1928–1940, 2008.
- [42] A. T. Frank, A. C. Stelzer, H. M. Al, and I. Andricioaei, “Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition,” *Nucleic Acids Research*, vol. 37, no. 11, pp. 3670–3679, 2009.
- [43] C. Musselman, Q. Zhang, H. Al-Hashimi, and I. Andricioaei, “Referencing strategy for the direct comparison of nuclear magnetic resonance and molecular dynamics motional parameters in RNA,” *Journal of Physical Chemistry B*, vol. 114, no. 2, pp. 929–939, 2010.
- [44] Q. Zhang, X. Sun, E. D. Watt, and H. M. Al-Hashimi, “Resolving the motional modes that code for RNA adaptation,” *Science*, vol. 311, no. 5761, pp. 653–656, 2006.
- [45] E. J. d’Auvergne and P. R. Gooley, “The use of model selection in the model-free analysis of protein dynamics,” *Journal of Biomolecular NMR*, vol. 25, no. 1, pp. 25–39, 2003.
- [46] G. Schwarz, “Estimating dimension of a model,” *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [47] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Proceedings of the 2nd International Symposium on Information Theory*, pp. 267–281, Akadémiai Kiadó, Budapest, Hungary, 1973.
- [48] A. M. Mandel, M. Akke, and A. G. Palmer, “Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme,” *Journal of Molecular Biology*, vol. 246, no. 6, pp. 144–163, 1995.
- [49] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, and W. F. van Gunsteren, “Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme,” *Journal of Biomolecular NMR*, vol. 30, no. 4, pp. 407–422, 2004.
- [50] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, “Comparison of multiple amber force fields and development of improved protein backbone parameters,” *Proteins*, vol. 65, no. 3, pp. 712–725, 2006.
- [51] V. Wong and D. Case, “Evaluating rotational diffusion from protein MD simulations,” *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6013–6024, 2008.
- [52] K. Lindorff-Larsen, S. Piana, K. Palmo et al., “Improved side-chain torsion potentials for the Amber ff99SB protein force field,” *Proteins*, vol. 78, no. 8, pp. 1950–1958, 2010.
- [53] S. A. Showalter and R. Brüschweiler, “Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field,” *Journal of Chemical Theory and Computation*, vol. 3, no. 3, pp. 961–975, 2007.
- [54] B. Zagrovic and W. F. van Gunsteren, “Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us?” *Proteins*, vol. 63, no. 1, pp. 210–218, 2006.
- [55] R. B. Best, N.-V. Buchete, and G. Hummer, “Are current molecular dynamics force fields too helical?” *Biophysical Journal*, vol. 95, no. 1, pp. L07–L09, 2008.
- [56] L. Wickstrom, A. Okur, and C. Simmerling, “Evaluating the performance of the FF99SB force field based on NMR scalar coupling data,” *Biophysical Journal*, vol. 97, no. 3, pp. 853–856, 2009.
- [57] D. W. Li and R. Brüschweiler, “NMR-based protein potentials,” *Angewandte Chemie*, vol. 49, no. 38, pp. 6778–6780, 2010.
- [58] D. W. Li and R. Brüschweiler, “Certification of molecular dynamics trajectories with NMR chemical shifts,” *Journal of Physical Chemistry Letters*, vol. 1, no. 1, pp. 246–248, 2010.
- [59] S. A. Showalter, E. Johnson, M. Rance, and R. Brüschweiler, “Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations,” *Journal of the American Chemical Society*, vol. 129, no. 46, pp. 14146–14147, 2007.
- [60] N. Trbovic, B. Kim, R. A. Friesner, and A. G. Palmer, “Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation,” *Proteins*, vol. 71, no. 2, pp. 684–694, 2008.

- [61] S. Genheden, C. Diehl, M. Akke, and U. Ryde, "Starting-condition dependence of order parameters derived from molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 6, no. 7, pp. 2176–2190, 2010.
- [62] O. F. Lange, D. van der Spoel, and B. L. de Groot, "Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR Data," *Biophysical Journal*, vol. 99, no. 2, pp. 647–655, 2010.
- [63] J. Tropp, "Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: the effect of fluctuating internuclear distances," *The Journal of Chemical Physics*, vol. 72, no. 11, pp. 6035–6043, 1980.
- [64] D. I. Freedberg, R. M. Venable, A. Rossi, T. E. Bull, and R. W. Pastor, "Discriminating the helical forms of peptides by NMR and molecular dynamics simulation," *Journal of the American Chemical Society*, vol. 126, no. 33, pp. 10478–10484, 2004.
- [65] A. D. Mackerell Jr., M. Feig, and C. L. Brooks 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 simulation," *Journal of Computational Chemistry*, vol. 25, no. 11, pp. 1400–1415, 2004.
- [66] B. Guillot, "A reappraisal of what we have learnt during three decades of computer simulations on water," *Journal of Molecular Liquids*, vol. 101, no. 1–3, pp. 219–260, 2002.
- [67] E. L. Ulrich, H. Akutsu, J. F. Doreleijers et al., "Biological Magnetic Resonance Bank," *Nucleic Acids Research*, vol. 36, no. 1, pp. D402–D408, 2008.
- [68] D.-W. Li and R. Brüschweiler, "Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins," *Journal of Chemical Theory and Computation*, vol. 7, no. 6, pp. 1773–1782, 2011.
- [69] K. Wüthrich, "Protein structure determination in solution by NMR spectroscopy," *Journal of Biological Chemistry*, vol. 265, no. 36, pp. 22059–22062, 1990.
- [70] P. Güntert, "Structure calculation of biological macromolecules from NMR data," *Quarterly Reviews of Biophysics*, vol. 31, no. 2, pp. 145–237, 1998.
- [71] P. Güntert, W. Braun, and K. Wüthrich, "Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA," *Journal of Molecular Biology*, vol. 217, no. 3, pp. 517–530, 1991.
- [72] P. Güntert, C. Mumenthaler, and K. Wüthrich, "Torsion angle dynamics for NMR structure calculation with the new program DYANA," *Journal of Molecular Biology*, vol. 273, no. 1, pp. 283–298, 1997.
- [73] T. Herrmann, P. Güntert, and K. Wüthrich, "Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA," *Journal of Molecular Biology*, vol. 319, no. 1, pp. 209–227, 2002.
- [74] B. López-Méndez and P. Güntert, "Automated protein structure determination from NMR spectra," *Journal of the American Chemical Society*, vol. 128, no. 40, pp. 13112–13122, 2006.
- [75] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore, "The Xplor-NIH NMR molecular structure determination package," *Journal of Magnetic Resonance*, vol. 160, no. 1, pp. 65–73, 2003.
- [76] C. D. Schwieters, J. J. Kuszewski, and G. Marius Clore, "Using Xplor-NIH for NMR molecular structure determination," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 48, no. 1, pp. 47–62, 2006.
- [77] G. M. Clore, W. L. Delano, P. Gros et al., "Crystallography & NMR system: a new software suite for macromolecular structure determination," *Acta Crystallographica D*, vol. 54, no. 5, pp. 905–921, 1998.
- [78] A. T. Brünger, "Version 1.2 of the Crystallography and NMR system," *Nature protocols*, vol. 2, no. 11, pp. 2728–2733, 2007.
- [79] M. Nilges, M. J. MacLus, S. I. O'Donoghue, and H. Oschkinat, "Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β -spectrin," *Journal of Molecular Biology*, vol. 269, no. 3, pp. 408–422, 1997.
- [80] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges, "ARIA2: automated NOE assignment and data integration in NMR structure calculation," *Bioinformatics*, vol. 23, no. 3, pp. 381–382, 2007.
- [81] K. T. Simons, R. Bonneau, I. Ruczinski, and D. Baker, "Ab initio protein structure prediction of CASP III targets using ROSETTA," *Proteins*, vol. 37, no. 3, pp. 171–176, 1999.
- [82] Y. Shen, O. Lange, F. Delaglio et al., "Consistent blind protein structure generation from NMR chemical shift data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 12, pp. 4685–4690, 2008.
- [83] Y. Shen, R. Vernon, D. Baker, and A. Bax, "De novo protein structure generation from incomplete chemical shift assignments," *Journal of Biomolecular NMR*, vol. 43, no. 2, pp. 63–78, 2009.
- [84] S. Raman, Y. J. Huang, B. Mao et al., "Accurate automated protein NMR structure determination using unassigned NOESY data," *Journal of the American Chemical Society*, vol. 132, no. 1, pp. 202–207, 2010.
- [85] S. Raman, O. F. Lange, P. Rossi et al., "NMR structure determination for larger proteins using backbone-only data," *Science*, vol. 327, no. 5968, pp. 1014–1018, 2010.
- [86] A. D. J. van Dijk, R. Boelens, and A. M. J. J. Bonvin, "Data-driven docking for the study of biomolecular complexes," *FEBS Journal*, vol. 272, no. 2, pp. 293–312, 2005.
- [87] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin, "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information," *Journal of the American Chemical Society*, vol. 125, no. 7, pp. 1731–1737, 2003.
- [88] S. J. de Vries, A. D. J. van Dijk, M. Krzeminski et al., "HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets," *Proteins*, vol. 69, no. 4, pp. 726–733, 2007.
- [89] A. van Dijk, D. Fushman, and A. Bonvin, "Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked di-ubiquitin and validation against ^{15}N -relaxation data," *Proteins*, vol. 60, no. 3, pp. 367–381, 2005.
- [90] A. D. J. van Dijk, R. Kaptein, R. Boelens, and A. M. J. J. Bonvin, "Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking," *Journal of Biomolecular NMR*, vol. 34, no. 4, pp. 237–244, 2006.
- [91] C. Schmitz and A. M. J. J. Bonvin, "Protein-protein HADDOCK using exclusively pseudocontact shifts," *Journal of Biomolecular NMR*, vol. 50, no. 3, pp. 263–266, 2011.
- [92] X. J. Morelli, P. N. Palma, F. Guerlesquin, and A. C. Rigby, "A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data," *Protein Science*, vol. 10, no. 10, pp. 2131–2137, 2001.
- [93] X. Morelli and A. C. Rigby, "Acceleration of the drug discovery process: a combinatorial approach using NMR spectroscopy and virtual screening," *Current Computer-Aided Drug Design*, vol. 3, no. 1, pp. 33–49, 2007.

- [94] M. R. Jensen, L. Salmon, G. Nodet, and M. Blackledge, "Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts," *Journal of the American Chemical Society*, vol. 132, no. 4, pp. 1270–1272, 2010.
- [95] J. A. Marsh and J. D. Forman-Kay, "Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints," *Journal of Molecular Biology*, vol. 391, no. 2, pp. 359–374, 2009.
- [96] M. D. Mukrasch, P. Markwick, J. Biernat et al., "Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation," *Journal of the American Chemical Society*, vol. 129, no. 16, pp. 5235–5243, 2007.
- [97] P. Bernadó, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge, "Defining long-range order and local disorder in native α -synuclein using residual dipolar couplings," *Journal of the American Chemical Society*, vol. 127, no. 51, pp. 17968–17969, 2005.
- [98] M. D. Mukrasch, S. Bibow, J. Korukottu et al., "Structural polymorphism of 441-residue tau at single residue resolution," *Plos Biology*, vol. 7, no. 2, Article ID e1000034, 2009.
- [99] D. Ganguly and J. Chen, "Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states," *Journal of Molecular Biology*, vol. 390, no. 3, pp. 467–477, 2009.
- [100] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, and M. Blackledge, "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 47, pp. 17002–17007, 2005.
- [101] A. K. Jha, A. Colubri, K. F. Freed, and T. R. Sosnick, "Statistical coil model of the unfolded state: resolving the reconciliation problem," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 37, pp. 13099–13104, 2005.
- [102] W.-Y. Choy and J. D. Forman-Kay, "Calculation of ensembles of structures representing the unfolded state of an SH3 domain," *Journal of Molecular Biology*, vol. 308, no. 5, pp. 1011–1032, 2001.
- [103] J. A. Marsh, C. Neale, F. E. Jack et al., "Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure," *Journal of Molecular Biology*, vol. 367, no. 5, pp. 1494–1510, 2007.
- [104] I. Bertini, Y. K. Gupta, C. Luchinat et al., "Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains," *Journal of the American Chemical Society*, vol. 129, no. 42, pp. 12786–12794, 2007.
- [105] I. Bertini, A. Giachetti, C. Luchinat et al., "Conformational space of flexible biological macromolecules from average data," *Journal of the American Chemical Society*, vol. 132, no. 38, pp. 13553–13558, 2010.
- [106] G. Saab-Rincón, P. J. Gualfetti, and C. R. Matthews, "Mutagenic and thermodynamic analyses of residual structure in the α subunit of tryptophan synthase," *Biochemistry*, vol. 35, no. 6, pp. 1988–1994, 1996.
- [107] P. Garcia, L. Serrano, D. Durand, M. Rico, and M. Bruix, "NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation," *Protein Science*, vol. 10, no. 6, pp. 1100–1112, 2001.
- [108] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw et al., "Long-range interactions within a nonnative protein," *Science*, vol. 295, no. 5560, pp. 1719–1722, 2002.
- [109] H. Tafer, S. Hiller, C. Hilty, C. Fernández, and K. Wüthrich, "Nonrandom structure in the urea-unfolded *Escherichia coli* outer membrane protein X (OmpX)," *Biochemistry*, vol. 43, no. 4, pp. 860–869, 2004.
- [110] D. A. C. Beck, A. L. Jonsson, R. D. Schaeffer et al., "Dynameomics: mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations," *Protein Engineering, Design and Selection*, vol. 21, no. 6, pp. 353–368, 2008.
- [111] M. W. van der Kamp, R. D. Schaeffer, A. L. Jonsson et al., "Dynameomics: a comprehensive database of protein dynamics," *Structure*, vol. 18, no. 4, pp. 423–435, 2010.
- [112] N. P. Cowieson, B. Kobe, and J. L. Martin, "United we stand: combining structural methods," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 617–622, 2008.
- [113] X. Wang, H.-W. Lee, Y. Liu, and J. H. Prestegard, "Structural NMR of protein oligomers using hybrid methods," *Journal of Structural Biology*, vol. 173, no. 3, pp. 515–529, 2011.