

How does it really move? Recent progress in the investigation of protein nanosecond dynamics by NMR and Simulation

Olof Stenström, Candide Champion, Marc Lehner, Guillaume Bouvignies, Sereina Riniker, Fabien Ferrage

▶ To cite this version:

Olof Stenström, Candide Champion, Marc Lehner, Guillaume Bouvignies, Sereina Riniker, et al.. How does it really move? Recent progress in the investigation of protein nanosecond dynamics by NMR and Simulation. Current Opinion in Structural Biology, 2022, 77, pp.102459. 10.1016/j.sbi.2022.102459. hal-03864085

HAL Id: hal-03864085

https://hal.science/hal-03864085

Submitted on 21 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How does it really move? Recent progress in the investigation of protein nanosecond dynamics by NMR and Simulation

Olof Stenström¹, Candide Champion¹, Marc Lehner¹, Guillaume Bouvignies¹, Sereina Riniker¹, Fabien Ferrage¹,

^aLaboratoire des biomolécules, LBM, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France ^bLaboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

Abstract

Nuclear magnetic resonance (NMR) spin relaxation experiments currently probe molecular motions on timescales from picoseconds to nanoseconds. The detailed interpretation of these motions in atomic detail benefits from complementarity with the results from molecular dynamics (MD) simulations. In this mini-review, we describe the recent developments in experimental techniques to study the backbone dynamics from ¹⁵N relaxation and side-chain dynamics from ¹³C relaxation, discuss the different analysis approaches from model-free to dynamics detectors, and highlight the many ways that NMR relaxation experiments and MD simulations can be used together to improve the interpretation and gain insights into protein dynamics.

^{*}Corresponding authors

Introduction

Experimental methods to determine the conformations of proteins and nucleic acids have empowered structural biologists to build a vast body of knowledge with tens of thousands of structures of proteins, nucleic acids, and their complexes. Protein structures have had a tremendous impact in our chemical understanding of protein function. Yet, any protein function involves protein motions, from the selective catalysis of a chemical reaction to signalling through interactions with binding partners. Such motions can often be described as a re-weighting of distributions of conformations that preexist functional events. Protein function is thus intimately dependent on the kinetics and energetics of interconversion between protein conformations. Here, we review recent progress in the experimental and computational investigation of internal protein motions, with a particular focus on nanosecond timescales.

Most experimental techniques employed to characterize the conformation of biological macromolecules, can be exploited to determine conformational ensembles: cryogenic electron microscopy [?]; (room-temperature) X-ray crystallography [?], nuclear magnetic resonance (NMR) [?], or small-angle X-ray scattering [?]. Yet, time-resolved X-ray crystallography and NMR are unique in their abilities to investigate dynamics at high-resolution over extremely broad ranges of timescales: from picoseconds (ps) to days with NMR [?] and femtoseconds (fs) to hours with X-ray crystallography [?]. In order to be concise, we consider in this review exclusively NMR approaches, which can tackle solutions, microcrystals, fibrils, and aggregates. We will particularly focus on the interplay between NMR relaxation exper-

iments and molecular dynamics (MD) simulations. We first describe recent developments in relaxation experiments and approaches to analyse relaxation rates and to interpret the molecular motion, followed by an in-depth discussion of the crosstalk between NMR and MD, and the connection to conformational entropy.

Recent Experimental Developments

Current Techniques

NMR relaxation originates from the fluctuations of interactions of nuclear spins (Figure ??). In biomolecules, these interactions are dipole-dipole couplings with other nuclei and sometimes with electrons, the anisotropic part of chemical shift tensors, as well as quadrupolar couplings (only for nuclei with a spin number of 1 or more, such as deuterons). These interactions are orientation-dependent so that their fluctuations originate predominantly from the rotation component of all motions: from methyl-group rotation, wobbling of a flexible loop in a protein, and, if in solution, overall rotational diffusion, or, if in a crystal, rocking motions. Correlation times for overall rotational diffusion in biomolecules amenable to solution-state NMR range from a few nanoseconds (ns) to a few tens of ns, which sets an upper bound on the timescale of internal motions accessible to relaxation in solution. NMR relaxation happens most often on millisecond to second timescales but originates from the fluctuations of interactions due to motions order of magnitude faster, in the ps and ns ranges. The nuclei most commonly used for relaxation studies are ¹⁵N for the protein backbone and bases in nucleic acids, ²H in methyl groups, and ¹³C in protein side-chains and nucleic acids. The

methods to measure longitudinal (R_1) , transverse (R_2) relaxation rates and dipolar cross-relaxation rates σ have been well established for about 20 years.

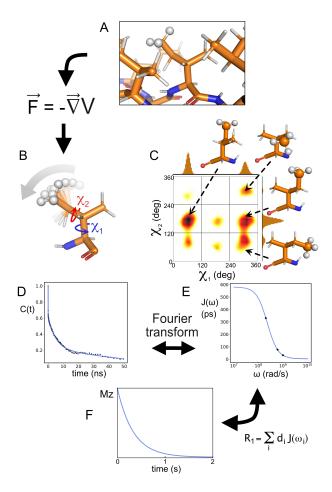


Figure 1: Protein motions from MD simulations and NMR. (a) MD simulations associate energies and forces to a given conformation that lead to (b) motions of atoms and groups of atoms. (c) When the simulation is long enough, the ensemble of accessible conformations is spanned, as illustrated here for residue isoleucine 13 in the protein ubiquitin. (d) Correlation functions for bond vectors C(t) can be computed from MD trajectories. (e) The Fourier transform of the correlation function is the spectral density function $J(\omega)$. (f) NMR relaxation rates are linear combinations of the spectral density function at the resonance frequencies of the spin system.

Extensive measurements covering most of the range of accessible commercial high magnetic fields have led to a better characterization of the spectral density function (see below), particularly in intrinsically disordered proteins [? ?], yet, only a narrow range of frequencies can be probed at high fields (from 9.4 to 28.2 T). In order to extend this range by orders of magnitude, Redfield introduced high-resolution relaxometry (HRR) [?] (Figure ??). In HRR experiments, the sample is transferred to a position of the stray field of the superconducting magnet for the relaxation delay. Thus, it is possible to measure site-specific longitudinal relaxation rates in proteins at magnetic fields as low as 0.3 T [? ?]. Both backbone dynamics from ¹⁵N relaxation [?] and side-chain dynamics from ¹³C relaxation in the specifically-labeled ¹³C¹H²H₂ methyl groups [?] have been characterized by HRR. Recent HRR measurements were analyzed with a protocol that takes into account spin dynamics during the transfer of the sample between high and low fields [?]. This approach was subsequently validated by relaxation measurements on a two-field spectrometer [?]. In all cases, motions of the backbone and sidechains in the range of hundreds of ps to a few ns have been identified. This highlights the additional information on slower modes of motions provided by relaxation at moderate magnetic fields (0.1 to 10 T). Identifying slower motions will require to reach lower magnetic fields but also somehow slowing down overall rotational diffusion.

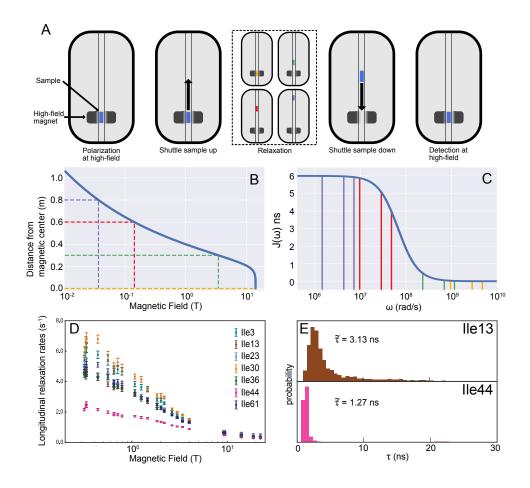


Figure 2: Principle of high-resolution relaxometry (HRR). (a) A generalized scheme of HRR: After polarization at a high magnetic field, the sample is transferred to a specific position in the stray field for relaxation. This is followed by shuttling down into the high field for detection. Four experiments with different transfer distances are represented by the colors yellow, green, red and purple with traveling distance of 0, 0.3, 0.6 and 0.8 m, respectively. (b) The magnetic field profile as a function of the distance from the magnetic center with the four distances from the magnetic center. (c) The spectral density of a carbon-13 nucleus, coupled to a single proton in a protein with a correlation time of 15 ns with the frequencies sampled by longitudinal relaxation of the carbon-13 nucleus at the four magnetic fields. (d) Magnetic-field dependent longitudinal relaxation rates of the 13 C nucleus in specifically labeled 13 C 1 H 2 H $_{2}$ $\delta 1$ methyl groups of isoleucine residues in the protein ubiquitin. (e) Probability distribution of the correlation time for ns motions in residues isoleucine 13 and 44 in ubiquitin 6 median values are shown.

How to Reach Longer Timescales?

As mentioned in the previous section, commonly used solution-state relaxation-based experiments are insensitive to motions slower than the molecular rotational tumbling correlation time, which generally lies in the low ns range. On the other hand, chemical exchange-based techniques, such as chemical-exchange saturation transfer (CEST), $R_{1\rho}$ and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments, are most sensitive to dynamic processes on the microsecond (μ s) to second timescale. Thus, the accurate description of the intermediate motional regime between the ns and the μ s timescale range remains a technical challenge.

Extending the domain of applicability of standard relaxation experiments to slower timescales can be achieved by artificially slowing down overall rotational tumbling, e.g., by increasing the solvent viscosity with viscogens such as ethylene glycol or glycerol [? ?], or by elongating the molecule of interest [?]. Overall rotational correlation times of a few tens of ns could then be obtained using these techniques.

Recently, Brüschweiler and co-workers introduced an elegant strategy, exploiting aqueous colloidal dispersions of synthetic nanoparticles to which a biomolecule can bind [?]. Provided that the exchange rate is fast on the chemical shift timescale ($> 10^3 \text{ s}^{-1}$) and slow on the molecular tumbling timescale ($< 10^7 \text{ to } 10^8 \text{ s}^{-1}$), the observed increase in site-specific transverse relaxation rates (ΔR_2) due to the binding to the nanoparticle is, to a good approximation, directly proportional to the order parameter (S^2) reporting on all the internal motions ranging from the ps up to the rotational tumbling correlation time of the nanoparticle. The latter can extend into the hundreds

of ns to μ s range depending on the size of the nanoparticle. This approach reveals internal motions over timescales two orders of magnitude slower than what standard relaxation approaches can achieve using only two samples, one with and one without nanoparticles.

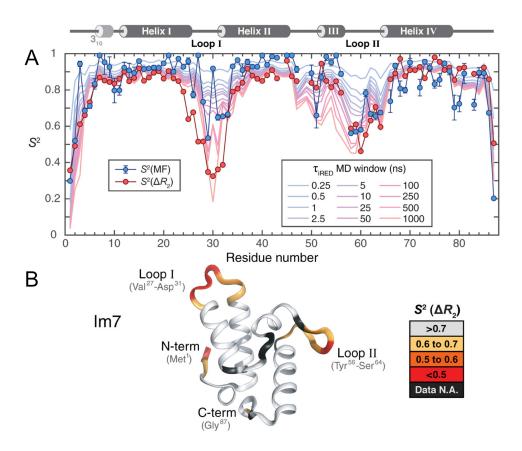


Figure 3: Nanosecond motions of a loop revealed by nanoparticle assisted spin relaxation. (A) Comparison of order parameters S^2 derived from model-free analysis of high-field nitrogen-15 relaxation (MF, blue circles) and NASR (ΔR_2 , red circles) in the protein Im7. Order parameters derived from MD simulations with variable averaging time window, from 250 ps to 1 μ s (solid lines). (B) Structure of Im7 color-coded with the order parameters from NASR analysis. Adapted from Xie et al. [?] with permission from AAAS (pending).

The nanoparticle-assisted spin relaxation (NASR) method was first applied to characterize the backbone motions of small globular proteins (Im7, CBD1 and ubiquitin). Here, 15 N ΔR_2 values were compared to traditional model-free (MF) order parameter values (S_{MF}^2) that report on ps to ns motions (Figure ??) [? ?]. Backbone S^2 profiles were remarkably similar on most parts of Im7 and CBD1 between the two approaches, lower order parameters in the NASR profiles were observed in a functionally important loop of Im7, revealing large-scale motions in the ns to μ s range [?]. On the contrary, no additional dynamics in the slower timescales were observed in ubiquitin [?]. More recently, the NASR method was used to probe side-chain dynamics in proteins using 2 H-relaxation of CH₂D methyl moieties and cross-correlated relaxation of CH₃ groups [?]. Additional side-chain dynamics in the ns to μ s timescale were found in functional loops of Im7 while side-chain motions appeared largely confined into the sub-ns range in ubiquitin, in agreement with the results from the NASR studies on backbone dynamics.

Overall rotational diffusion is usually absent in the solid state, opening up the full ns and μ s ranges. Over the past decade, many approaches have been introduced to probe the amplitudes of local motions from the scaling of anisotropic interactions: dipole-dipole coupling, chemical shift anisotropy, and quadrupolar couplings [? ? ? ? ?]. These approaches benefit greatly from specific isotope labeling, in particular in protein side chains [? ?], where uniformly labeled samples suffer from an over-abundance of dipole-dipole couplings. Work from the early 1980s [?] demonstrated that timescales of motions in solids could be determined from the analysis of longitudinal relaxation rates. The measurement of site-specific relaxation rates in solids

and their analysis in terms of local dynamics was introduced in the early 2000s [?]. Motions on timescales of tens of ns have been identified recently for the backbone of an amyloid fiber [?] and aromatic ring flips [?]. Another key advantage of relaxation in solids is the ability to explore a broad range of temperatures. The measurement of relaxation rates and quadrupolar interactions between about 100 K and 300 K in solid-state samples provides access not only to correlation times but also to the respective activation energies [??].

Analysis Approaches

Model-Free Approach

The value of experimental measurements to probe molecular motions lies in the quantitative information that can be determined from their analysis. NMR relaxation datasets rarely contain enough information to determine the precise geometries and timescales of all motions occurring over the ps and ns ranges. In particular, most investigations of protein motions from relaxation rely on measurements at only one or two magnetic fields. Analyzing such datasets requires simple models. In this context, the model-free approach (MFA) by Lipari and Szabo [?] has made it possible to obtain simple, quantitative information on the amplitude of ps to ns motions. The MFA approximates the correlation function for interactions (often internuclear vectors) to the product between correlation functions for internal motions and for overall rotational diffusion. This decoupling approximation is rigorous in the presence of isotropic rotational diffusion and relies on some hypotheses when the overall motion is anisotropic (i.e., timescale separation, statistical

independence, or symmetry of local motions) [?]. The internal correlation function is approximated to a simple mono-exponential decay with an effective correlation time and a limit at long times called the generalized order parameter S^2 . The success of the MFA lies in its simplicity. Yet, MFA provides no information on the nature, geometry, or distribution of timescales of motions. Several recent works have attempted to go beyond the MFA to describe internal motions in biomolecules in more detail.

Dynamic Detector Approach

Characterising macromolecular dynamics using the MFA can diverge significantly from the actual motion, as shown by Smith et al. initially for solid-state [? ?] and later for solution-state NMR [?]. If the internal motion is described by a tri-exponential correlation function and the corresponding relaxation rates are analysed using two exponentials, the resulting correlation times and amplitudes of motions can deviate significantly from the real motion. By using so-called dynamic detectors (DDs), it is possible to avoid the problem of incorrect modes of motions. DDs characterise the motion on a range of timescales, in contrast to the MFA that uses discrete correlation times.

The DD approach is based on the calculation of experimental sensitivities corresponding to the relaxation rates measured. These sensitivities report on the relaxation induced by motion as a function of correlation time. By taking linear combinations of the experimental sensitivities, it is possible to construct DDs, which are non-zero only in a limited range of correlation times and are optimized to cover different ranges. Applying the same linear combinations as for the optimized DDs on the experimental relaxation rates

gives the responses for each detector. The DD response reports on motion in the window of correlation times defined by that detector. This approach of a linear relationship between the distribution of motion and the relaxation rate is especially useful in solid-state NMR where no global tumbling is present and the correlation function principally describes internal motion. A similar DD methodology has been developed for solution-state NMR [?], taking into account the overall tumbling as well as the contribution to the transverse relaxation rate from chemical exchange processes.

The information obtained through the DD approach is less specific compared to information from more explicit models. However, comparing experimental DD responses directly with MD-derived responses can help to validate the simulation results and extract valuable insights [??]. Furthermore, the DD approach demonstrated that HRR can provide information on motions with correlation times in the ns range, which are not described as well solely using high-field relaxation data. However, the extent of available information depends on the global correlation time, and it is thus more useful to analyze macromolecules with moderate to slow tumbling [?].

Explicit Models

Ultimately, a goal of investigating molecular motions is to describe their nature: transitions pathways between conformations, geometric shapes of fluctuations, etc. Explicit models of motions predate the MFA and come from an era when developing a mathematically complex pen-and-paper model was more accessible than performing a long MD simulation [?]. Thus, rotation on a cone [?], wobbling in a cone [?], and rotamer jump [?] models were developed to interpret NMR relaxation measurements. For decades, limited

relaxation datasets have made it difficult to discriminate between models, so that these explicit models have been supplanted by the MFA. Today, broad relaxation datasets, such as those obtained from HRR measurements over orders of magnitude of magnetic field, make it possible to determine a larger number of parameters and discriminate models. In addition, MD simulations, although not perfect, are reliable enough (see below) to guide the development of explicit models of motions. Such models, combining rotamer jumps with rotation on a cone or Gaussian axial fluctuations (GAF), were developed to determine motions of side chains in peptides and proteins from relaxation measurements [? ?]. The combined analysis of MD simulations, HRR, and high-field measurements with explicit models was shown to better reproduce experimental data, opening the way for a mechanistic interpretation of motions from NMR relaxation [?].

NMR and MD Crosstalk

Comparison of Relaxation Rates

NMR relaxation experiments have been compared to MD simulations as early as the development of the MFA [?]. An overview and comparison of the different analysis methods is given schematically in Figure ??. Typically, the simulated and experimental data were both fit using a (extended) MFA, providing order parameters and correlation times (Figure ??). While this gives direct information about the timescale of specific motions, the interpretation can be biased by the choice of the model (e.g., number of exponential terms included) [?]. Common improvements include the four-parameter extended MFA [?], which splits the motion into overall tumbling and a fast and slow

internal motion. However, this also introduces more parameters to be determined, which can lead to over-fitting, and often requires further knowledge of the involved relaxation motions. Introduced in the mid-1990s, the spectral density mapping approach [?] has become popular in the MD community, and was shown to reproduce order parameters more accurately than direct fitting of correlation functions in a MFA [?].

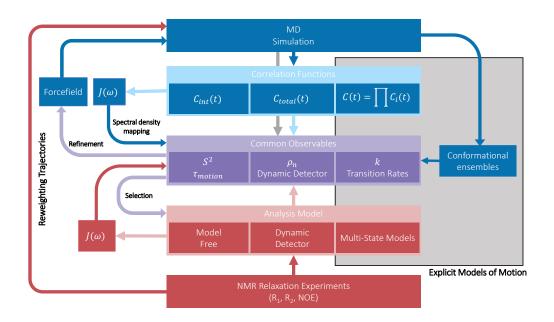


Figure 4: Schematic overview of the interplay between NMR relaxation experiments and MD simulations. Different analysis approaches may be used to reach a common observable which allows to ascertain the validity of the MD trajectory. Discrepencies in observables may also be used to guide forcefield development, or to reweight trajectories.

An alternative technique is the isotropic reorientational eigenmode dynamics (iRED) [?] method in which order parameters are obtained from a prin-

cical component analysis of the isotropically averaged covariance matrix of the dipolar interactions. Finally, the DD approach may also be used to compare with MD data by employing the same set of detectors optimized from experimental relaxation rates [? ?].

Many approaches comparing the motions from MD and experiment rely on the splitting of the total motion into an internal motion (which is fitted) and an overall tumbling, which is assumed to be isotropic and to average out. However, for molecules with large disordered parts, this is generally not the case and various degrees of anisotropic tumbling occurs [?]. Approaches like iRED can deal with anisotropic tumbling by calculating the correlation function as an average over spherical harmonics. Other methods like the DD approach can be adapted to include anisotropic tumbling by processing the trajectory in small frames of reference focused around different types of motion [?].

Enabled by advances in computing power, longer MD simulations have revealed that μ s simulations were required to obtain converged order parameters describing ps to ns motion [?]. Convergence can be improved by splitting a simulation into blocks, performing the analysis of each block and averaging the results [??]. If the results from MD and NMR disagree, Salvi et al. [?] have proposed an approach termed ABSURD (average block selection using relaxation data) to reweight blocks of simulation and reach a better agreement. To prevent over-fitting, Kümmerer and co-workers included an additional entropy contribution to the cost function in the reweighting procedure of ABSURDer [?].

Given a match between simulation and experiment, the wealth of infor-

mation contained in the MD trajectories may be further analyzed (e.g., with Markov state models [????]), or combined with additional experimental techniques [?].

Using NMR to Refine MD Force Fields

The quality of the underlying force field (FF) will directly affect the level of agreement between simulation and experiment. Biomolecular FFs have typically been fitted against thermodynamic data (e.g., density, heat of vaporization, and solvation free energies), ensuring a correct distribution of accessible configurational states. Recently, increased computing capabilities and comparison with NMR relaxation data have allowed to assess the validity of existing FFs to reproduce also kinetic data. Significant differences were observed among various FF families [? ?], pointing to specific parts of FFs that are in need of refinement. One avenue to improve FFs is to tune them using NMR observables (e.g., chemical shifts, 3J -coupling constants, S^2 order parameters) [? ? ?]. For example, small changes in the torsional parameters of methyl-bearing side chains led to improved agreement with experimental relaxation rates [? ?].

Special Case: Intrinsically Disordered Proteins

Implications for Experimental Methods

Intrinsically disordered proteins (IDPs) can be investigated by conventional NMR relaxation methods. However, the interpretation of relaxation rates in IDPs cannot rely on the separation of independent overall rotational diffusion and internal motions. This necessary paradigm shift in the analysis

of relaxation in IDPs has led several research groups to record extensive sets of relaxation rates [? ? ?], in order to identify differences in the spectral density function between IDPs and folded proteins. Such large sets of high-field relaxation rates were analyzed with models including three correlation times [? ?], or, alternatively discretized [?] and continuous [?] distributions. The information of high-field NMR relaxation alone is insufficient to decisively discriminate between these models but the combination of relaxation and MD simulations provides essential insight.

Implications for Simulations

In principle, IDPs should not be different from folded proteins regarding the FF. However, due to parametrization strategies used for biomolecular FFs (especially for backbone torsions) [?], an overstabilization of secondary (and tertiary) structure motifs has been observed with many FFs [???]. To address this, IDP specific FFs have been developed over the past few years [? ? ? ? ?]. The general strategy is to adjust the backbone torsions residue-specifically using the coiled regions of protein crystal structures as reference. Brüschweiler and co-workers [?] took thereby a hybrid approach for ff99SBnmr2 by taking the NMR-optimized ff99SBnmr1 [?] as basis. For a99SB-disp, Robustelli et al. [?] modified not only the torsional parameters but also the protein and water van der Waals interactions. The resulting FFs are applicable to both IDPs and folded proteins. These developments clearly show how sensitive slow protein dynamics are to small changes in the FF parameters and the choice of the water model. In the presence of imperfect FFs for IDPs in solution, NMR relaxation rates have been used to guide MD simulations in order to investigate the types of motions in IDPs that lead to

relaxation [? ?]: fast librations of the peptide plane on ps timescales, jumps of backbone torsions on timescales ~ 1 ns, and chain-like motions on effective timescales of several ns.

What About Entropy?

NMR order parameters report on the equilibrium distribution of orientations for a bond vector in the molecular reference frame. Changes in order parameter, due to ligand binding for example, can reflect changes of conformational entropy, defined as the entropy obtained if each accessible conformation is considered as an accessible microstate [? ? ?]. Recent studies have suggested the use of an entropy meter, which estimates the total conformational entropy difference through the average of order parameters from methyl bearing side-chains, with the assumptions that methyl groups gives good enough a coverage of the macromolecule and sufficient coupling to the surrounding side chains to report on local motions [? ?]. Hoffmann and co-workers [?] showed that most of the conformational entropy of the backbone is sampled within the correlation time of protein tumbling so that the order parameter contains information on most of the conformational entropy of the backbone. However, the lifetime of different sidechain rotamers can be longer than overall tumbling and the accurate distribution of rotamers will not be fully detected by conventional NMR relaxation experiments [? ?]. Thus, the conformational entropy was on average 30% higher when motions slower than overall tumbling were taken into account. These results indicate that a recalibration of the entropy meter is needed to access the total conformational entropy.

Conformational entropy of side chains can also be estimated from MD simulations by extracting the equilibrium populations of rotameric states from (sufficiently long) MD simulations.

Conclusions and Outlook

Describing motions on nanosecond timescales in proteins is a difficult challenge. MD simulations of proteins on μ s timescales are now routine, and the quantitative agreement of such simulations with experiments has been improving constantly. Yet, the resulting description of motions in proteins from ps to ns timescales is still very much improvable. On the experimental front, venturing in the ns range is a difficult objective for NMR relaxation methods as rotational diffusion in solution hides slower motions, and direct sampling of the spectral density at relevant frequencies is difficult at high magnetic fields. Recent progress with the transient slow down of rotational diffusion and the application of relaxometry to proteins offer still limited but quantitative information on ns motions. The most promising advances are now at the interface of relaxation and simulation methods, using the experimental information given by relaxation measurements to guide MD simulations directly or improve the post-processing of trajectories. As force fields have been mainly parametrized against thermodynamic data until now, their validation on kinetic data is crucial for future improvements. While most of the recent developments described here involved small to medium-size proteins, future advances will likely focus on larger proteins, where the benefit of new experimental and simulation methods will be more pronounced.

${\bf Acknowledgments}$

The authors gratefully acknowledge financial support by the European Union, Grant agreement 899683 (HIRES-MULTIDYN).