SAINT PETERSBURG STATE UNIVERSITY

Manuscript Copyright

Lebedenko Olga Olegovna

Calculations of measurable NMR parameters based on MD modeling data for biomolecular systems: new methods and applications

Scientific specialty
1.3.8. Condensed matter physics

Dissertation for the degree of
Candidate of Physical and Mathematical Sciences

Translation from Russian

Scientific Supervisor:
Ph.D.
Nikolai Ruslanovich Skrynnikov

Saint Petersburg — 2024

Table of contents

Pages.

Introduction

Relevance

After the first protein structure was determined by X-ray crystallography for myoglobin [1], protein molecules have been viewed as fully structured, static entities. It took a substantial amount of time until experimental evidence of conformational variability in proteins emerged. Among the first reports of protein dynamics, Wüthrich and Wagner have shown that side chains in the protein's hydrophobic core are capable of jumping between different rotameric states [2]. Thus, despite the existence of the uniquely defined structure, the hydrophobic protein core displays certain properties of a polymer melt. Later, the analyses of crystallographic data led another group of researchers to suggest that individual domains in multi-domain proteins can shift their positions relative to each other and that domain mobility plays an important role in enzymatic catalysis [3]. At around the same time, it has been shown that a number of protein molecules, including histone proteins, contain disordered fragments which, despite the absence of a well-defined structure, play an important functional role [4].

Currently, protein dynamics is a subject of active research. In some way or form, protein dynamics has an influence on protein folding and stability, ligand binding, allosteric effects in signal transduction and other functional properties of protein molecules. Intrinsically disordered proteins, as well as disordered regions in the otherwise structured proteins, play an important role in the architecture and mechanism of molecular engines, such as ribosome [5] in the spatial organization and function of chromatin [6] and nuclear pore complexes [7], in cytoskeletal assembly [8] and formation of phase-separated organelles [9]. While highly important, disordered peptide chains are vulnerable to proteolytic degradation, aberrant posttranslational modifications and, ultimately, to conversion into various pathological forms (including prions and amyloid fibrils). Due to these singular features, disordered proteins play a prominent role in neurodegenerative disease [10].

Probing disordered proteins and, more broadly, protein dynamics requires special experimental instruments, as well as new approaches toward modeling of

protein molecules, distinct from the traditional means of structure investigation. Structural studies mainly rely on X-ray diffraction data and, since recently, cryoelectron microscopy data. These data can shed some light on protein dynamics, but only indirectly and to a limited extent. For example, the absence of electron density while modeling a protein molecule guided by crystallographic or cryo-EM data suggests that the corresponding fragment of the peptide chain is highly mobile. To obtain a more detailed picture, one needs to employ a special experimental toolkit, dominated by Nuclear Magnetic Resonance (NMR) spectroscopy. NMR experiments make it possible to observe signals from the individual atoms in the disordered peptide chain and thus obtain information about the dynamic parameters pertaining to the near environment of a given atom, as well as its more distant environment. The measurements are typically based on multidimensional heteronuclear experiments that use ^{15}N and ^{13}C nuclei from protein backbone and side chains as probes of conformational dynamics. Using heteronuclei ^{15}N and ^{13}C as opposed to protons substantially simplifies the analyses since the results can be usually interpreted by considering a small-sized spin system, consisting of two-to-four spins.

Among many NMR observables, chemical shifts and spin relaxation rates have special significance. Chemical shift reflects the characteristics of the atomic close-range environment. In particular, chemical shifts are sensitive to the so-called "residual structure", i.e. the statistical propensity of a disordered sequence to form transient α-helices, β-hairpin or other, more complex, structural arrangements. Besides, analyses of NMR spectra allow one to identify the effect of the so-called "exchange broadening" caused by relatively slow (from microseconds to seconds) transitions between the conformational states with different chemical shifts. A significant number of specialized experiments have been developed to elucidate the parameters of such exchange dynamics. In turn, heteronuclear spin relaxation rates (primarily those of ^{15}N and ^{13}C spins) carry the information on amplitudes and characteristic time constants of certain motional modes that modulate spin interactions involving those nuclei. Typically, these modes are associated with reorientational dynamics of the vector that connects nitrogen or carbon atom with its bonded proton. The most popular relaxation measurements involve longitudinal (R_1) and transverse (R_2) spin relaxation rates. However, beyond that current experimental techniques are capable of measuring dozens of different relaxation parameters for various

multi-spin correlations (coherences). A special place in this battery of relaxation experiments belongs to the measurements of the so-called paramagnetic relaxation enhancement (PRE) effect. To measure PREs, a paramagnetic label is introduced into the system. Typically, it is a nitroxyl radical that is conjugated to a cysteine side chain by means of maleimide chemistry. The unique characteristic of the PRE effect is that it is a function of the long-range dipolar interaction between the paramagnetic center and the individual nuclear spin, which thus serves as a probe to characterize stochastic dynamics in the system at hand.

The repertoire of NMR methods to study disordered proteins is not limited to the analyses of chemical shifts and relaxation rates. Valuable information can also be obtained from NMR experiments using pulsed field gradients, which allow one to measure translational diffusion rates for disordered proteins. In turn, this makes it possible to characterize the protein's conformational ensemble in terms of overall compactness. The same information can also be gathered from small-angle X-ray scattering data, but this latter technique suffers from poor sensitivity when applied to small proteins or peptides. In addition, a number of other experimental methods are used to study disordered proteins, such as circular dichroism, analytical ultracentrifugation and proteolytic fragmentation. A separate class of experiments makes use of covalent labels: apart from the PRE measurements discussed above, this category includes Electron Spin Resonance (ESR), Förster resonance energy transfer (FRET) and a number of variations of the same theme (e.g. BRET), chemical cross-linking (e.g. disulfide trapping) and other methods. However, these techniques are less informative than the NMR-based experiments and the use of bulky labels can cause unwanted changes in the delicate conformational equilibria of disordered proteins.

Clearly, studying disordered proteins calls for a new type of models, different from static protein structures. Initially, disordered proteins were modeled using structural ensembles, i.e. sets of tens or hundreds of structures intended to reproduce the conformational diversity of a disordered protein. In these models, statistical weights have been assigned to the individual conformers reflecting their frequency of occurrence in the system at hand. A number of programs have been developed to assist with the construction of such ensemble models, most notably ENSEMBLE [11] and ASTEROIDS [12]. Briefly, their algorithm can be described as follows. As a starting point, the pool of many conformers is generated in a random manner (without any consideration for

conformational propensities of the given peptide sequence). Next, the statistical weights attached to all conformers are optimized such as to reproduce the set of experimental data for the protein in question. One of the shortcomings of such static models is the absence of temporal dimension, i.e. the information on characteristic transition rates for transitions from one conformational state to another. It should also be noted that the quality of such models is entirely dependent on experimental data that inform the selection procedure. Relative scarcity of the experimental data, the complicated character of data interpretation and low information content create a situation where construction of ensemble models turns into a major effort, while the resulting models prove to be poorly determined.

As an alternative solution, Xue and Skrynnikov proposed to use molecular dynamics (MD) trajectories as models of disordered proteins [13]. In the following years this method has been actively developed and applied to a wide range of disordered systems [14; 15]. In principle, MD-based models make it possible to fully characterize the behavior of a natively disordered protein, including the transition rates for interconversion between different conformational states. It should also be noted that MD-based models are a product of an a priori knowledge in a form of force field, which offers a fairly realistic description of the interactions between various peptide moieties, as well as the interaction of peptide with solvent water. Therefore, there are reasons to believe that reliable models of disordered proteins can thus be generated from the first principles without recourse to experimental data.

However, in practice this approach faces a number of technical hurdles. First, ultra-long MD simulations are needed to ensure adequate coverage of the vast conformational phase space accessible to a disordered protein. Advances in GPU computing, as well continuous improvements in MD algorithms (e.g. accelerated MD schemes) offer a workable solution to this problem, at least for disordered systems of smaller size. Second, biomolecular force fields have been traditionally designed, optimized and tested for well-structured globular proteins. For this class of systems a wealth of data has been amassed, making it possible to judge the strengths and weaknesses of the current force fields and revise them accordingly. The first attempts to model disordered proteins using a standard MD setup showed some significant systematic deviations from the experiment. As it turned out, conventional force fields tend to overestimate the propensity

of disordered proteins to form (unstable) α-helices as well as their tendency to coalesce, i.e. form compact molten-globule-like species [16]. Further studies have shown that these artefacts are caused by overestimation of the protein-protein interactions at the expense of protein-water interactions. To correct this problem, new variants of force fields have been developed, including new water models [17]. Yet subsequent tests have shown that some of these new models are poorly balanced, causing partial destabilization of globular proteins [18].

Thus, the situation developed where reliability of MD models of disordered proteins (or proteins containing disordered regions) remains in doubt. The attempts to validate these models so far have lacked consistency and focus. Dozens of different proteins and peptides have been investigated using a number of different MD protocols. The experimental datasets used to validate MD models have been rather heterogeneous and typically included just a few measurements. The calculations conducted as a part of validation procedure often relied on simplified and even crude computational schemes. As a result, numerous publications in this area paint a rather discordant picture, with different research groups advocating different modeling schemes and oftentimes making contradictory claims.

This rather unsatisfactory situation provides strong motivation for this dissertation research. In our study we have developed and applied special MD modeling schemes to facilitate subsequent calculation of NMR parameters. We have also implemented new and improved algorithms to calculate NMR variables from the MD simulation data. Finally, we compared the results of our calculations with the experimentally measured values. This type of analysis paves the way for more accurate and reliable validation of the MD models at hand. In turn, those MD models that have been successfully validated, can be regarded as a uniquely complete source of information on the dynamics of disordered proteins and, more broadly, on protein dynamics in general.

Aims

In the first chapter, we discuss the development of methodology to calculate diffusion parameters of disordered proteins based on MD modeling data. The study has been performed on N-terminal fragment of histone H4 (N-H4) and made use of the relevant experimental data. We sought to develop a

protocol to generate a large amount of MD data by effectively employing the available computational resources. Special attention has been paid to testing of different water models in the context of our diffusion study. For the sake of generality, we have also investigated translational and rotational diffusion of a small globular protein ubiquitin. The main question posed in our study can be expressed as follows: is it possible to draw a meaningful conclusion about the degree of compactness of the disordered protein's conformational ensemble based on the analyses of its diffusion properties? We also sought to compare the rigorous results of our MD-based treatment to various empirical schemes that have been developed to predict diffusion parameters from MD modeling data.

The second chapter is dedicated to calculations and interpretation of the paramagnetic relaxation rates (PRE) in a multicomponent system, which includes both structured and disordered elements. Specifically, the study was performed on a nucleosome core particle, where nitroxyl label has been attached to different sites on the surface of histone H3 (within its structured portion), while the spectral signals have been acquired from the flexible N-terminal region of histone H4. It has been clear to us that the PRE data collected from such samples should carry certain information about the localization of the H4 tail relative to the body of the nucleosome particle. Yet a number of questions remained. Is it possible to meaningfully interpret the PRE data based on the MD simulation data? In doing so, is it necessary to implement a rigorous scheme to calculate the PRE rates, taking into consideration translational dynamics of the mobile H4 tail, or is it sufficient to use the simplified Solomon-Bloembergen formula? Is it possible to achieve a desirable level of convergence in the PRE calculations based on MD trajectories with net length of several microseconds?

In the third chapter, we discuss the use of MD data to assist in interpretation of the experimental results pertaining to flips of phenylalanine side chains in the crystals of small globular protein ubiquitin. In this case, we were interested if MD modeling can be used as an additional source of information that complements the NMR results and, more specifically, if MD data can help to identify the origin of the splitting that has been observed for some of the spectral resonances, to explain the absence of certain signals, and to clarify the nature of conformational transitions in phenylalanine side chains. Besides, we also sought to identify structural factors that influence the flip rate of phenyl rings and are responsible for large differences in the experimentally

observed rates.

Scientific and practical significance

1. As described above, the studies of disordered proteins are faced with the deficit of experimental information. In this situation, measurements of translational diffusion coefficient by means of the NMR experiments using pulsed field gradients becomes a potentially important piece of the study, allowing for characterization of the protein's conformational ensemble in terms of its compactness / extendedness. In this dissertation, we present the scheme for MD modeling of a disordered protein (peptide) that permits highly rigorous calculation of its diffusion parameters. As a first step, the trajectory of a protein is recorded in a small-sized simulation cell. Next, the snapshots from this trajectory are used as starting points to record a set of short trajectories in the medium- and large-sized cells. Extrapolating the results to the limit of an infinitely large cell yields the desired value of the diffusion coefficient. This approach makes it possible to efficiently use all available computational resources since the series of short trajectories can be recorded using a cluster of computational servers under the management of a queueing system. Furthermore, this approach is conducive to MD modeling in the NVE ensemble (which is the most rigorous MD modality) since short trajectories do not suffer from any appreciable temperature drift. In our work we have also shown that the diffusion of disordered proteins can be successfully modeled not only with the standard integration step of 1-2 fs, but also with the extended step of 4 fs. To this end, it is necessary to apply hydrogen mass repartitioning (HMR) scheme to the protein molecule as well as solvent water. This result opens up further opportunities to accelerate the computations.

Modeling and calculations have been conducted on the disordered N-terminal fragment of histone H4 (N-H4), which consists of 25 amino-acid residues. Our study of peptide's diffusion has shown that the results are significantly influenced by viscosity of MD water. Hence, a systematic assessment of water viscosity for TIP4P-Ew, TIP4P-D and OPC models has been undertaken as a part of this study. After correcting for solvent viscosity, we found that the resulting diffusion coefficient for the N-H4 peptide immersed in TIP4P-Ew water is significantly overestimated compared to the experimental result. This is a

consequence of exceedingly compact nature of N-H4 ensemble in the TIP4P-Ew simulations. On the contrary, the diffusion parameters derived from TIP4P-D and OPC simulations are in good agreement with the experimental data. Thus, we conclude that TIP4P-Ew model is ill-suited for modeling of disordered proteins, whereas the two other water models can be legitimately used in such studies. This conclusion is supported by our analyses of relaxation data for the backbone ^{15}N spins in the N-H4 peptide. Therefore, our work provides an example of MD models successfully validated using the diffusion and spin relaxation data. Choosing between the two successful models, TIP4P-D and OPC, one may prefer the latter option as it requires only minimal correction with regard to viscosity and produces somewhat more accurate predictions for the diffusion rate of the small globular protein ubiquitin.

In the current literature one can find numerous examples of papers where translational diffusion data are used to validate various MD models of disordered proteins (peptides). However, instead of the rigorous D_{tr} calculations, such as demonstrated in this dissertation, the authors resort to various empirical tools, e.g. programs for calculation of hydrodynamics parameters HYDROPRO and HullRadSAS, as well as Kirkwood-Riseman relationship and phenomenological formula by Nygaard. Comparing the results from these empirical treatments with the outcome of the high-level analysis implemented in our study, we observe that the empirical methods tend to perform poorly. In fact, their results can be entirely misleading, e.g. incorrectly suggest that TIP4P-Ew is the optimal choice of solvent to model a disordered peptide. This happens because the programs HYDROPRO and HullRadSAS, as well as the empirical formulas that approximate their results, are intended for use on static (rigid) protein structures and, therefore, fail to take into consideration the effect of segmental dynamics and its influence on translational diffusion of disordered proteins.

2. Nucleosome core particle (NCP) is a fundamental structural unit of genetic information storage system in the cell. NCP is comprised of eight histone proteins that are assembled into a spool, with 1.7 turns of double-stranded DNA wound around this spool. The distinctive feature of histones is the presence of extended terminal fragments (tails) that are conformationally disordered and projected into solvent. Histone tails are of major functional importance, offering binding sites for transcription factors and many other chromatin-associated proteins, and also having a role in chromatin assembly. The interaction of histone

tails with their numerous ligands is regulated by means of a special "histone code" that involves covalent modifications of selected side chains. Due to their disordered nature, histone tails cannot be studied by means of the standard structure elucidation methods, leaving NMR spectroscopy as essentially the only experimental tool to probe their conformation and dynamics. Among many available NMR experiments, PRE measurements are of special significance since they can be used to explore the localization of histone tails relative to the body of the nucleosome particle. To measure the PRE effect, paramagnetic labels are attached to the structured portion of the NCP (in our study, residues 36, 65, 79 and 125 in histone H3 have been chosen as labeling sites), whereas the signals are registered from the residues in the mobile histone tail (in our case, the N-terminal tail of histone H4 isotopically enriched with ^{15}N). Of note, until recently there has been no theoretical model to accurately interpret the PRE data in such multicomponent system involving both structured and disordered elements.

As a part of this dissertation work, we have explored the possibility to interpret the PRE data in nucleosome particle based on the conformational ensemble model supplemented with the well-known Solomon-Bloembergen equation. Our study has found that this interpretation is essentially baseless. For example, the experimental PRE data can be reproduced using the model that consists of mere two conformers, which is obviously inconsistent with the general notion of histone tails dynamics. Similarly, the data can be reproduced using a set of 42 conformers, where the two conformers determine the PRE values while the remaining ones are essentially random. This unsatisfactory situation is a consequence of the extremely high sensitivity of PRE rates to the distances separating the paramagnetic center from $^{1}H^{N}$ nuclear spin probes.

To address this situation, we proposed an alternative interpretation which is based on the first-principles calculations of the PRE rates from the MD data. These rigorous calculations account not only for reorientational dynamics of the vector that connects paramagnetic center with the nuclear spin, but also for modulation of the interspin distance due to the translational motion of the mobile histone tail. The results from the NCP simulation in TIP4P-D water performed by the author, as well as the series of trajectories in the OPC water recorded by Peng and co-workers are in good qualitative agreement with the experimental data. Thus, in this work we have successfully validated the MD model of nucleosome particle using the set of experimental PRE data. As a next

step, the so verified MD model has been used to glean the information about the spatial positioning of the H4 tail. Our analyses showed that the N-terminal tail of histone H4 is localized mainly in the vicinity of the nucleosomal DNA, which can be attributed to the electrostatic interaction between the positively charged tail and the negatively charged DNA. At the same time, the tail retains its high mobility while engaged in "fuzzy" interaction with the DNA. The localization of the H4 tail in the vicinity of the DNA limits its accessibility to chromatin-associated proteins, including the histone-modifying enzymes and transcription factors.

To obtain further insight into dynamics of mobile histone tails in the nucleosome particle, it is necessary to improve the convergence of the MD-based PRE calculations. As demonstrated in our study, even the trajectories with the length of tens of microseconds do not guarantee an acceptable convergence. The main reason for this deficiency is the extreme sensitivity of the PRE rates to the interspin distances, as already pointed out above. To improve the convergence, we recommend the use of distributed computational resources, accelerated modeling schemes (e.g. involving 4 fs integration step) and other similar solutions.

3. Side-chain jumps in a globular protein can be viewed as an indicator of protein plasticity, suggesting the presence of significant dynamic fluctuations in the protein structure. Our co-workers from the research group of P. Schanda investigated the dynamics of phenylalanine side chains in three different crystalline forms of ubiquitin. The measurements were conducted by means of the solid-state NMR spectroscopy under magic angle spinning (MAS) conditions using the new isotopic labeling scheme where ^{1}H and ^{13}C spins are selectively incorporated into ε_1, ε_2 positions of the phenyl ring. Both phenylalanine residues in ubiquitin, Phe4 and Phe45, are located on the surface of the protein molecule, which makes them potentially sensitive to the effect of crystal contacts. In this dissertation work we have shown that the analyses of MD simulation data can supplement and enhance the interpretation that is originally based on the experimental NMR results.

In particular, we have shown that the splitting of the Phe4 signal in the cubic crystal form is caused by the stacking of the two Phe4 rings at the crystallization interface (where both residues belong to chains B in the asymmetric unit). By processing the MD modeling data with the program

SHIFTX2 we have calculated the chemical shifts of $^{1}H^{\varepsilon}$ and $^{13}C^{\varepsilon}$ spins, including ring-current contributions from the stacked phenyl rings, thus corroborating the assignment of the spectral signals in this crystal form. Our analyses of the MD data also indicate that dynamics of the phenylalanine side chains is dominated by ε_2 jumps and, hence, can be described as phenyl ring flips, while the torsional angle χ_1 remains essentially constant. MD modeling data further suggest that ring flips in Phe4 occur on the time scale of nanoseconds, whereas in Phe45 the respective time scale is in the range of tens or hundreds of microseconds; of note, this relationship holds for all three crystal forms investigated in this study. This result explains the absence of the experimentally observed signals from residue Phe45, which can be attributed to the effect of exchange broadening caused by microsecond dynamics. As a part of our MD study, we have also attempted to rationalize the observed difference in the flip rates of Phe4 and Phe45. In particular, we hypothesized that the flip rates depend on the side-chain conformation, the locally available volume (which can be characterized via solvent accessible surface area of the phenylalanine residue), or the crystal contacts. None of these hypotheses, however, found direct support in the MD data, suggesting that the flip rate is instead determined by an interplay of more subtle cooperative effects within the ubiquitin structure. Thus, it has been demonstrated that MD simulations of protein crystals is a valuable source of information, complementing the results of the experimental study by solid-state NMR spectroscopy.

Conferences participation

The elements of this dissertation study have been presented at the following scientific meetings:

1. «65-th Biophysical Society Annual Meeting», online, 22 February – 26 February 2021.
2. International conference «Modern development of magnetic resonance 2021», Kazan, Russia, 1 November – 5 November 2021.
3. VI annual conference of the Translational Biomedicine Institute at SPbU «Topical problems of translational biomedicine – 2022», St. Petersburg, Russia, 25 July – 25 July 2022.

4. «Chinese Biophysics Congress 2024», Lanzhou, China, 25 July – 28 July 2024.

Author contribution

Chapter 1. The author has been personally responsible for planning of the research, recording, processing and analyzing the MD data, interpreting the results and preparing materials for publication. In particular, the author has devised and implemented a new scheme of MD modeling to quantify translational diffusion of disordered proteins. The experimental NMR data employed in this study have been mainly obtained by V.A. Salikov working jointly and under the supervision of I.S. Podkorytov in the Laboratory of Biomolecular NMR at SPbU. While processing and analyzing the MD data, the author used a collection of scripts originally developed by S.A. Izmailov.

Chapter 2. The author has been personally responsible for planning and implementing the part of this study related to the MD modeling, processing and analyses of the MD data, interpretation of the results and preparing the relevant materials for publication. In particular, the author has implemented in a form of program code a number of theoretical schemes to compute paramagnetic relaxation (PRE) rates based on the MD simulation data. The preparation of the nucleosome samples and PRE measurements by means of NMR spectroscopic experiments have been performed by W. Sun in the laboratory of C. Jaroniec.

Chapter 3. The author has been personally responsible for planning and implementing the part of this study related to MD modeling, processing and analyses of the MD data, interpretation of the results and preparing the relevant materials for publication. The preparation of the samples of crystalline ubiquitin and the experimental study of internal dynamics in these samples have been performed by D. Gauto and L. Becker in the laboratory of P. Schanda. During all stages of this work, the research problems were proposed and the results were interpreted under the scientific supervision of N.R. Skrynnikov.

Publications

The main results of the dissertation study are reported in 3 papers published in the scientific journals indexed in Scopus and Web of Science databases:

1. Using NMR diffusion data to validate MD models of disordered proteins: Test case of N-terminal tail of histone H4 / O. O. Lebedenko [et al.] // Biophysical Journal. — 2024. — Jan. — Vol. 123, no. 1. — P. 80—100.

2. Conformational and Interaction Landscape of Histone H4 Tails in Nucleosomes Probed by Paramagnetic NMR Spectroscopy / W. Sun [et al.] // Journal of the American Chemical Society. — 2023. — Sept. — Vol. 145, no. 46. — P. 25478—25485.

3. Aromatic ring flips in differently packed ubiquitin protein crystals from MAS NMR and MD / D. F. Gauto [et al.] // Journal of Structural Biology: X. — 2023. — Dec. — Vol. 7. — P. 1—12.

The scope and structure of the work

The thesis consists of an introduction, three chapters describing the main results of the dissertation study, conclusion and reference section. In turn, each chapter includes an introduction, description of the obtained results and their discussion, description of the employed materials and research methods, conclusions and supplementary materials. The work is presented on 174 pages, including 37 figures and 9 tables. The reference section contains 310 literature references.

The main scientific results

1. A highly effective MD modeling scheme has been developed to predict diffusion parameters of disordered proteins [19].
2. It has been shown that the MD-derived translational diffusion parameters reflect (i) the viscosity of the modeled water and (ii) the degree of compactness of the modeled protein's conformational ensemble [19].

3. It has been shown that the empirical approaches to calculation of disordered proteins' diffusion rates using MD modeling data have intrinsically poor accuracy and frequently lead to qualitatively incorrect conclusions [19].
4. It has been found that MD modeling of translational diffusion can be accelerated by using 4 fs integration step, but in order to retain high accuracy of the model one should apply the hydrogen mass repartition (HMR) scheme not only to the protein molecule, but also to the molecules of solvent water [19].
5. A rigorous theoretical scheme has been implemented to calculate paramagnetic relaxation rates (PREs) based on MD modeling data for a system containing disordered regions (nucleosome particle) [20].
6. It has been shown that the MD model of nucleosome particle is in qualitative agreement with the experimental PRE data and can be used to reconstruct the spatial distribution of the flexible histone "tails" [20].
7. By analyzing the set of ultra-long MD trajectories, we have shown that PRE calculations are extremely demanding with respect to convergence of the MD simulations, whereas the attempts to build a model in a form of static conformational ensemble lead to ill-defined results [20].
8. It has been shown that MD models of protein crystals can supplement and enhance the results of dynamics studies relying on experimental NMR data and, in particular, allow one to assign some of the observed NMR signals and explain the absence of the other signals [21].

Summary notes to be presented at the defense

1. A new scheme for MD modeling of disordered proteins presented in this dissertation allows one to efficiently generate MD data that are needed for calculation of the protein's translational diffusion rate.
2. Comparative analyses of the experimental data by diffusion NMR and the results of the MD-based calculations makes it possible to assess the validity of the MD model with regard to the compactness of the IDP's conformational ensemble. For smaller peptides, this information is unique since the alternative technique, small-angle X-ray scattering, suffers in this case from large experimental errors.

3. The presented methodology to calculate diffusion parameters based on MD modeling data led us to conclude that the empirical algorithms, which are widely used for this purpose, suffer from low accuracy and often mischaracterize the conformational ensemble of the modeled protein.
4. In this dissertation study, we report a new scheme for calculation of paramagnetic relaxation rates based on the MD modeling data; the scheme rigorously accounts not only for reorientational dynamics of the molecular system, but also for segmental dynamics of its disordered elements (viz. mobile histone "tails" in the nucleosome particle).
5. The analyses of the experimental PRE data in conjunction with the MD model of nucleosome particle indicate that the positively charged histone tails are localized in the vicinity of the negatively charged nucleosomal DNA whereby it is engaged in "fuzzy" interaction with the DNA while retaining a relatively high degree of conformational mobility.
6. According to our analyses of the PRE data based on the MD model of nucleosome particle, in order to ensure good convergence of the PRE calculations one needs to record MD trajectories with the net length of hundreds of microseconds. The extreme sensitivity of the PRE data to interspin distances makes it impossible to construct reliable PRE-based structural models where the conformational diversity of histone tails is represented in a form of static conformational ensemble.
7. In the study of crystalline ubiquitin by means of solid-state NMR spectroscopy, the MD modeling assists in spectral assignment and interpretation of the experimental data on rotameric jumps of protein side chains.

Chapter 1. Diffusion of disordered proteins: combined NMR and MD study of the N-terminal tail of histone H4

1.1 Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) have gained much attention over the last three decades. From the evolutionary perspective, IDPs are considered to be an advanced element of cellular machinery: the fraction of disordered sequences increases from archeal proteomes to bacterial proteomes to eukaryotic proteomes [22]. IDPs or IDPRs play a prominent role in molecular machines such as ribosome [23], in chromatin [6], in nuclear pore transporters [7], in cytoskeletal assembly [8] and in phase-separated organelles [9]. In broad terms, the prominence of IDPs and IDPRs stems from their involvement in cell signaling. Signaling networks involving disordered proteins have a higher degree of interconnectedness and a broader range of dynamic responses, resulting in improved functional efficiency [24]. The increased efficiency, however, comes at a price: disordered proteins are susceptible to proteolytic cleavage [25], aberrant modifications [26], conversion to a prionic form [27] or misassembly such as formation of amyloid fibrils [28]. All of this explains the ominous role of IDPs in neurodegenerative disorders, as well as wide involvement in cancer [29].

A search for pharmaceuticals to target IDPs is ongoing. However, finding effective small-molecule ligands of IDPs (which is prerequisite for intracellular targets) remains an outstanding challenge [30]. Industry-wide efforts to develop therapeutic antibodies against extracellular IDPs such as Aβ and extracellular tau so far have proved futile [31; 32].

From a structural standpoint, IDPs can be thought of as dynamic ensembles consisting of a multitude of constantly interconverting conformers. Roughly speaking, an unbound IDP can be viewed as a «statistical coil». Typically, it acquires structure upon binding to its structured target, cf. «folding upon binding» [33], although in some cases IDPs remain partially disordered after forming the so-called «fuzzy interaction» [34; 35] or even fully disordered after an engagement with another IDP [36]. Historically, unbound IDPs were

assumed to be featureless coils that were of little interest *per se* [37]. However, eventually it has been realized that IDPs have certain structural propensities, termed «residual structure» [38; 39]. These structural propensities can have an effect on IDPs' affinities for their biological targets. The effect is usually modest, yet biologically significant. For example, the level of residual helicity in transactivation domain of p53 has an influence on its binding constant to Mdm2 [40]. This and many other such examples have spurred interest in structural studies of intrinsically disordered proteins.

Standard methods of structural biology such as X-ray crystallography, cryo-EM or NOE-based NMR are unsuitable for characterization of IDPs and IDPRs or, otherwise, provide only indirect information. Instead, the bulk of experimental evidence about structural propensities of IDPs comes from other NMR data, such as heteronuclear chemical shifts, heteronuclear relaxation rates and paramagnetic relaxation enhancements (PREs), as well as diffusion data obtained from pulsed field gradient NMR (PFG NMR) experiments. In particular, diffusion data offer a potentially important piece of information, shedding light on compactness of the IDP's conformational ensemble (which can have a different appearance ranging from a highly extended random coil to a rather densely packed molten globule). The PFG NMR measurements could be particularly useful for smallish IDPs (disordered peptides) with molecular weight of several kDa. For systems of this size it can be difficult to obtain comparable information by means of small-angle X-ray scattering (SAXS) since the sensitivity of SAXS declines at low molecular weight [41] and boosting the signal by using high-brilliance synchrotron sources often causes radiation damage to protein samples [42]. In contrast, diffusion NMR spectroscopy can be successfully used for IDPs of an arbitrary size.

While in the case of folded proteins the experimental data translate into structural models that are deposited into Protein Data Bank, what are the comparable models for disordered proteins? Early on, IDPs conformational ensembles have been modeled by a collection of multiple conformers with attached statistical weights. The best-known ensemble generation schemes are ENSEMBLE [11] and ASTEROIDS [12]. Later, it has been proposed that MD trajectories can serve as structural models of IDPs, allowing for direct testing against NMR observables and other experimental data [13]. Indeed, MD simulations offer a natural path toward a realistic IDP's conformational

ensemble. Moreover, they capture the dynamic aspect of IDP's conformational equilibrium, i.e. encode the transition rates between a multitude of different conformers. This is particularly valuable for analyses of relaxation rates and PREs, which depend on motional time constants — these observables can be calculated directly from the MD trajectories and then compared with the experimental data.

While the idea of using an MD trajectory as a «structural-dynamic model» of an IDP seemed appealing, it faced some major challenges. First, the conventional MD force fields (and water models) proved to be ill-suited for modeling of disordered proteins. Second, typical MD trajectories were too short to sample the vast conformational phase space of an IDP. Over the last decade, a substantial headway has been made to address both of these problems. A number of new force fields and water models have been developed to model unstructured (as well as structured) proteins [18; 43—45]. At the same time, the advent of GPU computing along with development of special schemes for enhanced conformational sampling improved the situation with statistics of IDP simulations [15; 46—48]. As a consequence, it became possible to generate a converged trajectory involving a short IDP (i.e. a long peptide) and use it as a *bona fide* model to test against the experimental data [49; 50]. Otherwise, for longer IDPs, Blackledge and co-workers have developed the ABSURD approach whereby the disordered protein is modeled using a collection of shorter MD trajectories, which enter into the model with adjustable weights [51].

In this report, we address the question of whether an MD model of an intrinsically disordered protein can be validated against PFG NMR diffusion data. As already mentioned, the translational diffusion coefficient D_{tr} is one of the potentially informative pieces of data, characterizing the overall compactness of the IDP's conformational ensemble. However, calculating D_{tr} from the MD data (for the purpose of further comparison with the experimental result) is less than trivial.

Indeed, it is well-known that calculated D_{tr} values depend on the size of the simulation box. This has been demonstrated for simple fluids, polymers and proteins [52—58]. The origin of this effect can be traced to the constraints imposed on the MD simulation and, more specifically, the requirement of zero net momentum. This requirement leads to a situation where some of the dynamics in the simulated system is subtracted out; the effect is significant for small-sized

systems, leading to slower-than-expected apparent diffusion [59]. In order to counter this problem, Yeh and Hummer proposed to conduct several simulations in boxes of increasing size and then extrapolate the resulting D_{tr} values toward the limit of an infinitely large box [59]. While this approach has been commonly accepted, it is rather demanding from a computational standpoint. In particular, it requires one to record a long IDP simulation in a very large water box.

Furthermore, accurate determination of D_{tr} under the commonly used NPT ensemble equipped with Langevin thermostat proves to be problematic. It is a general observation that simulations employing Langevin thermostat do not reproduce hydrodynamics [60]. For the typically used values of friction constant γ (that are necessary for effective temperature control), the friction leads to appreciable increases in solvent viscosity and, accordingly, results in underestimated D_{tr} [61]. In principle, empirical schemes can be developed to correct for this effect; e.g., Hicks and co-workers recently presented such corrections for rotational diffusion coefficient D_{rot} [62]. However, these corrections are rather cumbersome and depend on a size of the simulated protein; in the case of IDPs, this approach is likely to face further complications. A better solution to this problem is to use another thermostat, such as Bussi-Parrinello velocity rescaling thermostat [63].

In this study we have focused on the intrinsically disordered N-terminal fragment of human histone H4 (residues 1-25). The N-terminal tail of H4 (hereafter referred to as N-H4) plays an important role in regulating gene transcription as well as chromatin architecture and remodeling [64]. In addition to the N-H4, we have also studied the well-known globular protein ubiquitin (Ub), which serves as a point of reference in our study of translational diffusoin. Since the MD-predicted D_{tr} constants of N-H4 and Ub are dependent on the viscosity of the modeled solvent, we have also systematically investigated the self-diffusion of water [65].

The simulations were conducted in the program Amber20 [66] under ff14SB force field [67] using three different water models: TIP4P-Ew [68], TIP4P-D [18], and OPC [69]. For both Ub and N-H4, we observe that the simulations using TIP4P-Ew significantly overestimate the diffusion coefficients, the simulations using TIP4P-D somewhat underestimate them, while the simulations using OPC overestimate them but only very slightly. These observations are largely attributable to variations in viscosity between the three water models. However,

in the case of the disordered N-H4 peptide the results also convey some information about the conformational ensemble of the peptide. Specifically, our data suggest that TIP4P-Ew solvent leads to overly compact representation of the N-H4 species. This observation is confirmed by the comparison of MD-simulated and experimental ^{15}N relaxation data. On the other hand, both OPC and TIP4P-D simulations are compatible with the experimental results within the uncertainty range.

Interestingly, analyses of the MD data using HYDROPRO software [70] or HYDROPRO-based parameterization [71] leads to the opposite conclusion, i.e. that the TIP4P-Ew model is the best water model to simulate N-H4. This serves as a cautionary note regarding the application of HYDROPRO, which is neither intended nor suited to be used on disordered proteins. Similarly, caution should be exercised with regard to other empirical schemes to predict D_{tr} using the simulated conformational ensembles of IDPs [72; 73]. Direct determination of D_{tr} from MD simulations, as demonstrated in this work, can provide a useful benchmark for further studies in this area.

1.2 Results and discussion

1.2.1 Water models for IDP simulations

The field of biomolecular MD simulations is dominated by classical 3-site or 4-site rigid point-charge water models [74]. The interactions of these water models are limited to pairwise Coulomb attraction/repulsion and pairwise Lennard-Jones (LJ) attraction (dispersion)/repulsion. The geometry of the water molecule, the magnitudes of point charges and their placement, as well as the LJ constants are usually adjusted to reproduce certain experimental characteristics of water. For example, the widely used 3-site TIP3P model approximates the geometry of water molecule in gas phase, with tunable interaction parameters adjusted such as to reproduce water density and heat of vaporization [75]. The 4-site TIP4P model, reported by the same authors, has the negative charge slightly shifted away from the oxygen atom, which improves

the agreement with the oxygen-oxygen and oxygen-hydrogen radial distribution functions from the neutron scattering experiments. Later TIP4P model has been re-optimized by using the Ewald summation technique instead of simply truncating the Coulomb interactions [68]. The resulting TIP4P-Ew model along with the classical TIP3P and SPC/E [76] have been the most popular choices in the field of biomolecular MD simulations over the last two decades.

About fifteen years ago, emerging interest in IDPs and advent of GPU computing prompted a number of research groups to begin simulating flexible peptide molecules. Soon it has been recognized that such simulations suffer from inadequate solvation of peptide chains [16]. Specifically, solvation free energies of peptide moieties turned out to be less favorable than those measured experimentally. A number of attempts have been made to repair this problem by fine-tuning the parameters of the relevant van der Waals interactions [77; 78]. In particular, these efforts led to a new version of Amber ff99SB force field [17; 79], which was later dubbed ff99SB-UCB. This modified force field has been complemented with a suitably modified TIP4P-Ew water model. Specifically, van der Waals parameters of the oxygen atom were tuned to better reproduce solvation free energies for a set of small molecules representative of peptide chemical space. Unfortunately, this modified water model remains nameless.

Soon thereafter, another modified water model has been introduced under the name TIP4P-D [18]. This model was built upon TIP4P/2005 [80], with van der Waals potential originating on the oxygen atom reparametrized such as to make peptide hydration more enthalpically favorable. Charges were also adjusted to fit the temperature-dependent density and heat of vaporization data. The model has been tested with a number of force fields, including Amber ff99SB-ILDN [81], and showed good results in IDP simulations. However, the authors also noted a slight tendency of TIP4P-D to destabilize globular proteins.

Two years later, a similarly modified TIP3P model has been introduced for use in conjunction with the amended force field, CHARMM36m, with the intent to simulate both folded and unfolded proteins [43]. Likewise, this revised water model was constructed by redefining van der Waals parameters in the original model. Unfortunately, this water model is also lacking a unique name.

A similar effort has also been undertaken to generalize Amber force fields, with a99SB-disp aspiring to model both folded and unfolded proteins [44]. This force field was equipped with a slightly revised version of TIP4P-D, which also

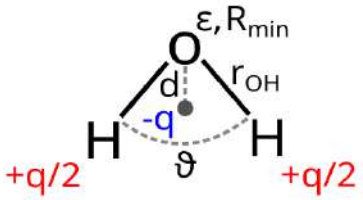
lacks a distinctive name. The new package was extensively tested on a set of 21 benchmark proteins.

In the meantime, a different type of water model, OPC, has been developed by optimizing positions of point charges with no regard for covalent geometry; the van der Waals parameters of the oxygen atom were tuned to match certain key experimental metrics [69]. Later, the authors reported that OPC model achieves good results in modeling of IDPs [45]. In particular, the trajectory of 26-residue N-terminal peptide from histone H4 recorded in OPC water under Amber99SB force field showed a significantly more expanded conformational ensemble compared to the trajectory in conventional TIP3P water.

A number of papers have also been published which compared the performance of the newly developed water models combined with various popular force fields [14; 49; 82—89]. The authors used radii of gyration measured by SAXS, chemical shifts and $^{3}J(H^{N},H^{\alpha})$-couplings measured by NMR, as well as a host of other diverse experimental characteristics, to test and validate the new water models. From these studies it emerged that the new water models clearly perform better in simulating disordered proteins than the previous-generation models such as TIP3P. However, beyond that it is difficult to draw any conclusions — it appears that some systems are better modeled by one water/force field combination, while other systems favor other combinations. It should also be added that only TIP4P-D model has been tested in more than a few studies, whereas other models received little attention. Altogether, the field needs more information on this problem, including a wider range of IDPs, more comprehensive experimental datasets, longer trajectories and improved study designs. The search for an optimal water model should be generally framed as a search for an optimal force field, including water model as one of its most consequential elements.

In this study we have chosen three water models, TIP4P-Ew, TIP4P-D and OPC, to investigate their performance in the context of MD modeling of protein diffusion, with the focus on disordered proteins. TIP4P-Ew is a classical model, which is selected here as a point of reference. TIP4P-D is a better established specialized model, developed for disordered proteins. Finally, OPC represents the latest generation of water models, aspiring to model both structured and disordered proteins. All of them are 4-site fixed-charge fixed-geometry models, which facilitates the comparison (e.g. there is no difference in the computational

costs between these models). For reader's convenience, the parameters of these models are summarized in Figure 1.1.



$$V_{LJ}^{water-water}(r) = \varepsilon\left[\left(\frac{R_{min}}{r}\right)^{12} - 2\left(\frac{R_{min}}{r}\right)^{6}\right]$$

TIP4P-EW	
q	1.0484
r_{OH}	0.9572 Å
ϑ	104.52
d	0.1250 Å
ε	0.1628 kcal/mol
$R_{min}/2$	1.7759 Å

TIP4P-D	
q	1.1600
r_{OH}	0.9572 Å
ϑ	104.52
d	0.1546 Å
ε	0.2238 kcal/mol
$R_{min}/2$	1.7763 Å

OPC	
q	1.3583
r_{OH}	0.8724 Å
ϑ	103.60
d	0.1594 Å
ε	0.2128 kcal/mol
$R_{min}/2$	1.7771 Å

Figure 1.1 — Parameterization of water models TIP4P-Ew, TIP4P-D and OPC. Van der Waals parameters are for the oxygen atom, with the variable representing oxygen-to-oxygen distance r.

To interpret our MD results on protein diffusion, which is the main focus of this chapter, we also need the data on viscosity of the modeled bulk water. To address this question, we consider self-diffusion coefficients D_{tr}^{H2O} for TIP4P-Ew, TIP4P-D and OPC water models. While the data on self-diffusion coefficients for various water models are available in the literature, they are often obtained from simulating a relatively small cluster of water molecules with no regard for box-size dependence of D_{tr} [65]. Furthermore, the published data do not always document the temperature dependence of D_{tr}^{H2O}. Therefore, we decided to quantify D_{tr}^{H2O} as a part of this study, using the same simulation setup as employed in our study of protein diffusion.

Specifically, for each of the considered water models we recorded three 150-ns-long trajectories involving small-, medium- and large-sized water boxes. The box dimensions were taken to be the same as in Ub simulations (for instance, the large OPC box contained 65243 water molecules). The simulations have been conducted at two temperatures, 298 and 303 K (relevant for N-H4 and Ub, respectively). The trajectories were processed as described in section 1.3, resulting in three D_{tr}^{H2O} values per water model per temperature. These results are plotted in Figure 1.2 as a function of the inverse box size, $1/L$ (color-coded as indicated in the legend). Linear extrapolation of D_{tr}^{H2O} to the limit of an infinitely

Table 1 — MD-pred-cted self-diffusion coefficients $D_{(tr,0)}^{H2O}$ for three different water models at 298 and 303 K together with the corresponding experimental results

Temperature (K)	D_{tr}^{H2O} (10^{-9} m^2/s)			
	Experiment	MD simulations		
		TIP4P-Ew	TIP4P-D	OPC
298	2.30	2.63	2.13	2.41
303	2.60	2.96	2.39	2.60

large box, $1/L = 0$, allows one to recover the true MD-predicted self-diffusion coefficients of water. These extrapolated values $D_{(tr,0)}^{H2O}$, corresponding to the y-intercepts of the dashed lines in the graph, can be compared with the relevant experimental data [90], as indicated by the horizontal red lines. The results are also summarized in Table 1.

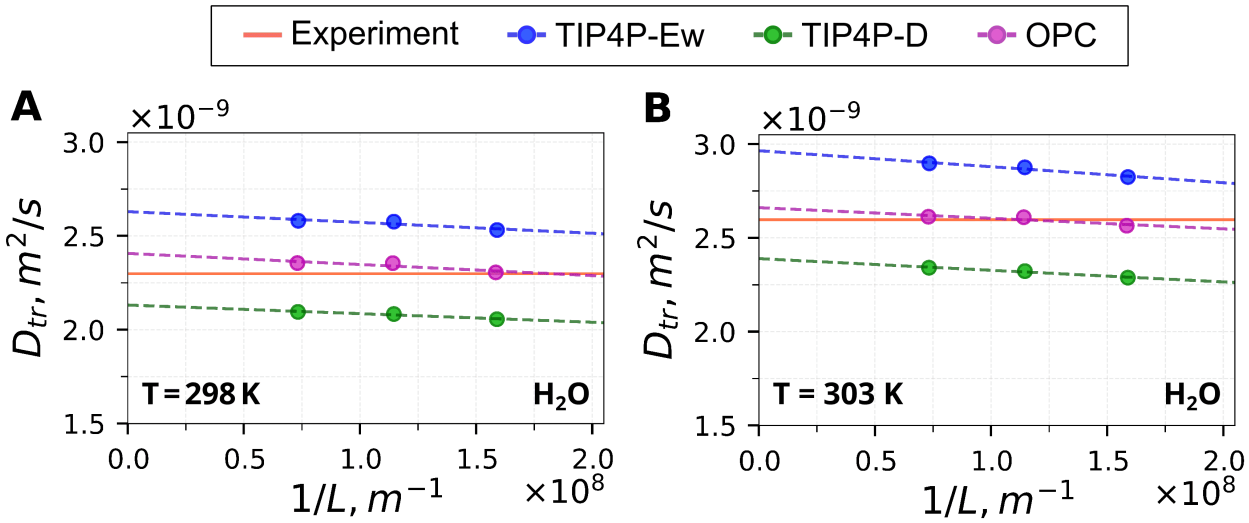


Figure 1.2 — Determination of self-diffusion coefficients D_{tr}^{H2O} from the series of MD simulations of water using TIP4P-Ew, TIP4P-D and OPC models at (A) 298 K and (B) 303 K. The results are plotted as a function of the inverse linear size of the simulation box, $1/L$. The per-box D_{tr}^{H2O} values are extrapolated to the limit of an infinitely large box, $1/L = 0$, using simple linear regression (dashed lines). The y-intercept of each regression line corresponds to the predicted self-diffusion coefficient for a given water model, $D_{(tr,0)}^{H2O}$ (listed in Table 1). The experimental results [90] are represented by solid red lines.

The inspection of Figure 1.2 confirms that there is slight, but distinct size dependence of the obtained D_{tr}^{H2O} values. Therefore, if the objective is

to compare the MD results with the experimental data one should use the extrapolated values, $D^{H2O}_{(tr,0)}$. Likewise, when discussing the properties of solvent water in the protein simulations, it is appropriate to refer to $D^{H2O}_{(tr,0)}$. Surveying the results in Figure 1.2 and Table 1, we observe that TIP4P-Ew model leads to appreciably overestimated $D^{H2O}_{(tr,0)}$, TIP4P-D model leads to somewhat underestimated $D^{H2O}_{(tr,0)}$, whereas the OPC model leads to minimally overestimated $D^{H2O}_{(tr,0)}$.

It is also instructive to compare the results with the information in the literature. For TIP4P-Ew water, the self-diffusion coefficient at 298 K was reported to be $2.4 \cdot 10^{-9}$ m^2/s [68], which is appreciably lower than the value of $2.63 \cdot 10^{-9}$ m^2/s reported here. This can be understood by noticing that the original estimates have been obtained using a small water cluster (512 molecules) simulated under the Andersen thermostat [91], which is, strictly speaking, not suitable for modeling of transport properties [61]. A similar observation concerning TIP4P-Ew has been recently made by the Economou group, who reported the size-corrected diffusion coefficient of $2.7 \cdot 10^{-9}$ m^2/s for this water model [65].

For the OPC water, the originally reported self-diffusion coefficient at 298 K was $2.3 \cdot 10^{-9}$ m^2/s [69], which is slightly lower than the value $2.41 \cdot 10^{-9}$ m^2/s found in our study. One should bear in mind that the water cluster simulated by Izadi et al. was also rather small (804 molecules) and the Langevin thermostat used by the authors of that study tend to bias the extracted diffusion rates (discussed in the section 1.1).

Finally, for the TIP4P-D water, the self-diffusion coefficient of $2.1 \cdot 10^{-9}$ m^2/s has been originally reported at the temperature of 300 K. This is slightly lower than the values $2.13 \cdot 10^{-9}$ and $2.39 \cdot 10^{-9}$ m^2/s which we obtained at 298 and 303 K, respectively. Of note, the authors used the 3054 water molecules and Nose-Hoover thermostat [91; 92], which is appropriate for this type of problem.

1.2.2 MD modelling of N-terminal tail of histone H4

We have recorded an extensive series of MD simulations for disordered N terminal tail of histone H4 (net length 75 μs, see Table 4 for details). This

positively charged peptide is comprised of the highly mobile glycine-rich segment (residues 1-15) and the segment that consists of bulkier residues with a very high proportion of charged amino acids (residues 16-25). There have been some early reports of α-helical propensity in N-H4, particularly in the acetylated form of the peptide, as well as propensity to adopt β-sheet conformations [93—96]. However, subsequent experimental studies and simulations of the isolated N-H4 peptide, as well as the N-terminal H4 tail within the nucleosome core particle (NCP) have not found any evidence of residual secondary structure in this segment [49; 97—102]. Of interest, NMR studies of NCP samples have exposed some intriguing aspects of N-H4 dynamics, including fuzzy interactions with the nucleosomal DNA (scalable by lysine acetylation) and a competition between H4 and H3 tails for the DNA interaction sites [50; 100—102].

Reviewing the trajectories of N-H4 in the TIP4P-Ew, TIP4P-D and OPC water, we note that in all of them the peptide dynamically interconverts between a multitude of random conformations. However, in the case of TIP4P-Ew simulation, it tends to constantly form various hairpin-like arrangements; sometimes, the entire peptide forms a semblance of hairpin, which continually morphs from one shape to another, but remains recognizable for up to several hundreds of nanoseconds. Occasionally, the peptide chain forms a «mini fold», which features little or no secondary structure, but likewise can retain its topology for up to several hundreds of nanoseconds. Such fluid structural motifs are held together by a handful of backbone-to-backbone, sidechain-to-backbone and sidechain-to-sidechain hydrogen bonds. In contrast, in the TIP4P-D trajectory the peptide mainly adopts extended conformations. Different hairpin- and loop-like motifs are more local in character and appear less frequently. Among them there is a distinctive small hairpin that is sporadically formed by the C terminal portion of the peptide [49]. Finally, the OPC trajectory falls somewhere in between the other two, as it features a mix of the extended conformations and loosely packed conformations.

From the perspective of diffusion measurements, it is essential that N-H4 adopts more compact conformations in the TIP4P-Ew solvent, more extended conformations in the TIP4P-D solvent, and a mixture of the two in the OPC solvent. This can be conveniently visualized through the gyration radius distribution function, $P(R_g)$. These distributions for the N-H4 simulations in the TIP4P-Ew, TIP4P-D and OPC water are shown in Figure 1.3 (blue, green and

magenta histograms, respectively). In each case, the distribution can be loosely described as bimodal. For the TIP4P-Ew histogram, the dominant narrow peak is centered just under 10 Å, with the secondary broader peak appearing at around 13 Å. For the TIP4P-D distribution, the broad main peak is centered at approximately 16 Å, while a minor component is visible at 11 Å. Finally, in the OPC histogram the broad peak with the larger area under the curve is positioned at ca. 15 Å, whereas a sharper peak with a roughly triangular shape occurs at 11 Å. The average R_g values are 11.5, 13.8 and 15.3 Å for TIP4P-Ew, OPC and TIP4P-D trajectories, respectively — indicating that the TIP4P-Ew and TIP4P-D results are wide apart, with OPC somewhat closer to the latter rather than the former.

The results shown in Figure 1.3 are expectable. As indicated previously, TIP4P-Ew favors intra-protein interactions at the expense of protein-to-water interactions, thus giving rise to (transiently) structured, more compact N-H4 species. At the same time, TIP4P-D shifts the balance in the opposite direction, emphasizing amply solvated, extended species. In the context of this work, it is important that the three water models, TIP4P-Ew, TIP4P-D and OPC predict significantly different conformational ensembles. The key question is whether the experimental D_{tr} measurements can help to discriminate between these three distinctive N-H4 ensembles. This question will be addressed in the remainder of the chapter.

To conclude this section, we will briefly discuss the situation with Ub simulations (see Table 5 for the summary of trajectories). Theoretically speaking, one can imagine that Ub molecule becomes somewhat «compacted» in the TIP4P-Ew solvent and somewhat «expanded» in the TIP4P-D solvent. This can conceivably happen due to the surface side chains that cling to the protein surface in the TIP4P-Ew solvent, but become extended outward in the TIP4P-D solvent. In fact, this kind of effect proved to be minor, with average R_g values of ubiquitin 11.69, 11.70 and 11.81 Å in TIP4P-Ew, OPC and TIP4P-D trajectories, respectively (small-box NPT simulations). Therefore, we can safely assume that, insofar as diffusion characteristics are concerned, the structural state of ubiquitin does not depend on the water model used in the simulations.

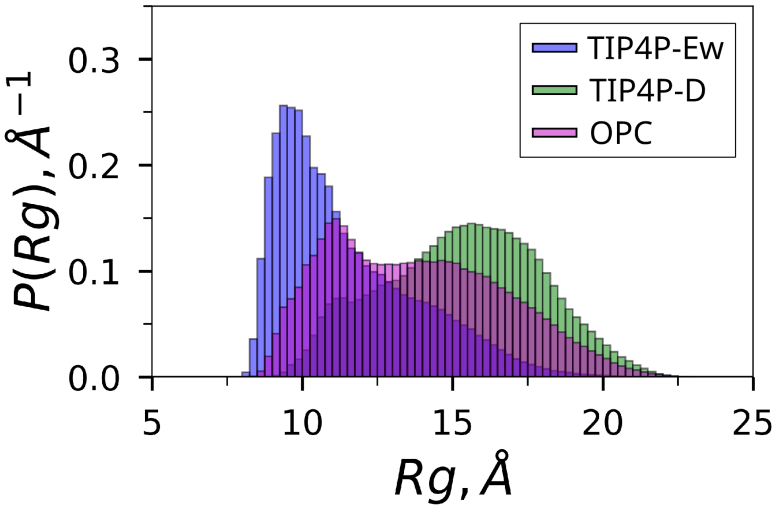


Figure 1.3 — Gyration radius distributions $P(R_g)$ for the N-H4 simulations in the TIP4P-Ew, TIP4P-D and OPC water (small-box trajectories, NPT simulations). The histograms for medium- and large-box NPT simulations, Figure 1.12 (B, C), are near-identical to the ones shown in this plot; this is not surprising given that the medium- and large-box trajectories were recorded as a set of short 10-ns segments starting from the frames that have been extracted from the small-box trajectory (see Figure 1.8). The histograms for small-, medium- and large-box NVE simulations employing TIP4P-Ew and TIP4P-D simulations, Figure 1.12, (D-F), are similar but not identical to the ones shown in this plot. Specifically, the distributions derived from NVE and NPT simulations in TIP4P-D solvent are, in fact, almost indistinguishable, whereas in the case of TIP4P-Ew there are some small, but visible differences. This outcome is understandable — indeed, one should expect that TIP4P-D simulations involving extended and highly dynamic peptide species are better converged than TIP4P-Ew simulations involving a host of interconverting hairpin-like conformers and other loosely structured arrangements.

1.2.3 Protein diffusion coefficients from the MD simulations

The MD data from the simulations of N-H4 in the TIP4P-Ew, TIP4P-D and OPC solvent have been processed as described in 1.3 to determine the box-size-dependent D_{tr}^{N-H4} quantities, which were then extrapolated to the limit of an infinitely large box, yielding the predictions for translational diffusion coefficients, $D_{(tr,0)}^{(N-H4)}$. The latter step is illustrated in Figure 1.4 (A) for the

N-H4 simulations in the NPT ensemble (the net duration of the trajectories 45 µs). The per-box values of D_{tr}^{N-H4} (circles in the plot, colored according to the type of water model used in the simulations) are successfully fitted with straight lines, thus confirming the sound character of the box-size correction scheme [59] and indicating that our results are well-converged. We further observe that the predicted diffusion coefficient $D_{(tr,0)}^{(N-H4)}$ from the simulations in the TIP4P-Ew water significantly overestimate the experimental result (cf. the y intercepts of the dashed blue line and the solid red line in Figure 1.4 (A)). At the same time, the simulations in TIP4P-D water (green symbols) and OPC water (magenta symbols) lead to moderate under- and overestimation of the experimental diffusion coefficient, respectively. The OPC result is particularly close to the experimental value, see Table 2.

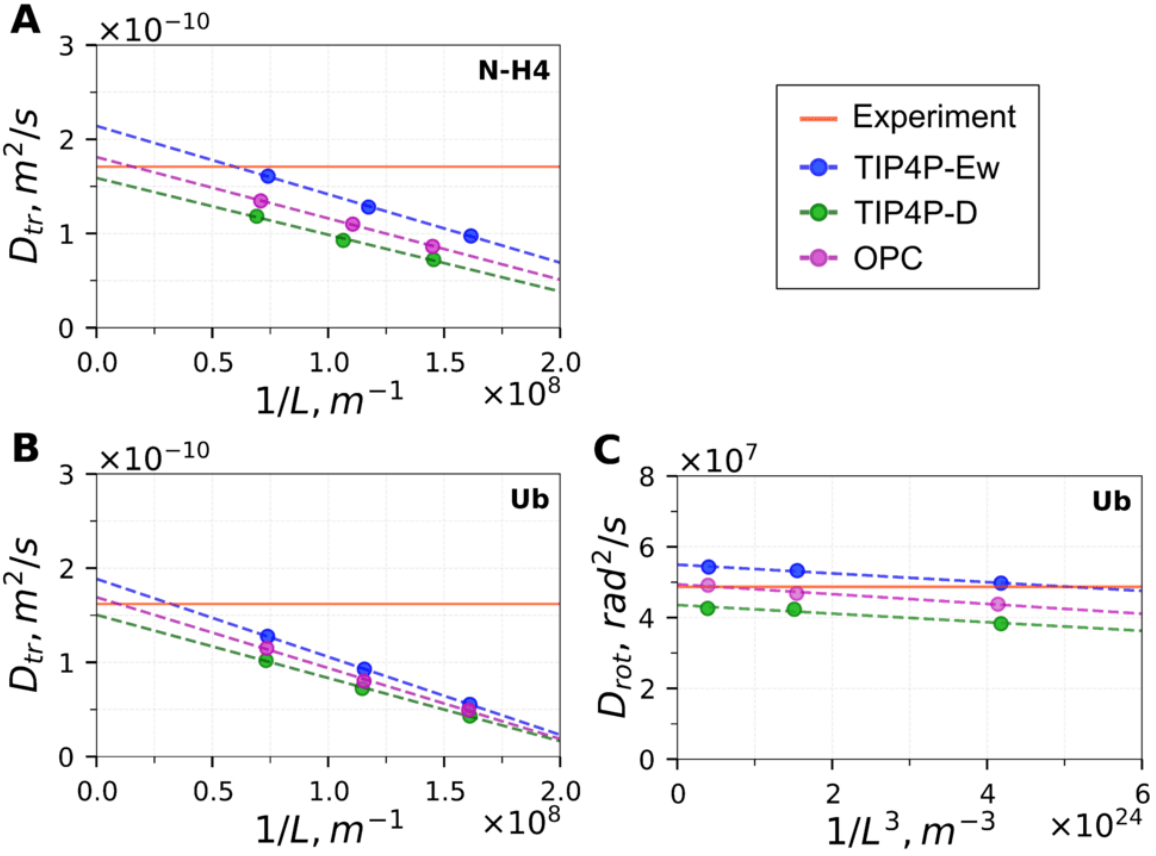


Figure 1.4 — Determination of diffusion coefficients for (A) N-H4 and (B, C) Ub from the series of MD simulations using TIP4P-Ew, TIP4P-D and OPC water models. The simulations have been conducted in the NPT ensemble at 298 K (for N-H4) and 303 K (for Ub). The per-box D_{tr} values are plotted as a function of the inverse linear size of the simulation box, $1/L$, and fitted via the linear regression (dashed lines). The y-intercept of each regression line corresponds to the predicted diffusion coefficients $D_{(tr,0)}^{(N-H4)}$ or $D_{(tr,0)}^{Ub}$. The per-box $D_{rot}^{U}b$

values are plotted as a function of the inverse (generalized) box volume, $1/L^3$, and likewise fitted via the linear regression. The y-intercept of each regression line corresponds to the predicted rotational diffusion coefficient $D^{Ub}_{(rot,0)}$. The positioning of the points in graphs B and C may lead one to assume that the y intercept of the D_{rot} vs. $1/L^3$ dependence is determined more reliably than that of the D_{tr} vs. $1/L$. However, there is some amount of scatter associated with D_{rot} values (readily visible for the middle point in panel C). Ultimately, the comparison of the results from the NPT and NVE simulations, as well as jackknife estimation of uncertainties (see Table 6) indicate that D_{tr} and D_{rot} have been determined with the similar precision. In addition to the simulated data, the experimental results are also shown in the plot (horizontal red lines).

The same kind of analysis has also been conducted for D^{Ub}_{tr} translational and D^{Ub}_{rot} rotational diffusion of Ub (the net duration of the NPT simulations 18 μs). The results shown in Figures 1.4 (B, C) display the same pattern as described above: TIP4P-Ew predictions are significantly overestimated, TIP4P-D predictions are underestimated, while OPC predictions are close to the target but slightly over the mark. This outcome can be explained by considering the following two factors. First, the viscosity of the simulated solvent varies between the different water models, see Figure 1.2 and Table 1. Second, in the case of N-H4, the simulated conformational ensemble of the peptide also varies between the models, see Figure 1.3. We will analyze these two factors and their impact on the predicted D_{tr} and D_{rot} values in the next section.

In addition to the NPT simulations, we have also generated a series of N-H4 and Ub trajectories using the NVE ensemble. These simulations have been conducted in TIP4P-Ew and TIP4P-D solvent, totaling 30 μs for N-H4 and 12 μs for Ub. The trajectories have been processed and analyzed in the same manner as discussed above. The availability of the independent D_{tr} and D_{rot} data obtained from the NVE simulations offers a good opportunity to test the convergence of our computational scheme.

A direct comparison of the diffusion coefficients obtained from the independent NPT and NVE simulations is presented in Figure 1.13 and also summarized in Table 6. The agreement is indeed very good, which confirms the suitability of the Bussi-Parrinello thermostat for modeling of molecular diffusion. The average (unsigned) deviation between the NPT and NVE results amounts to

mere 2.0 %. This is similar to the uncertainty of the experimental measurements — for example, the translational diffusion coefficient of Ub determined in this study has an uncertainty of 2.5 % (see Table 6). Therefore, we conclude that the precision of our calculations is sufficient to conduct a meaningful comparison to the experimental data. The calculated and experimental diffusion coefficients should be regarded as fully consistent so long as they fall within several percentage points of each other.

1.2.4 Comparison of experimental and MD-based results

In Figure 1.5 we visualize the deviations (expressed in percentage points) between the MD-predicted and experimental diffusion coefficients for water, Ub and N-H4. It is convenient to first discuss the case of Ub, where the conformational state of the protein is uniquely defined and protein diffusion is presumably controlled by solvent viscosity alone. The data on solvent viscosity in the form of D_{tr}^{H2O} at the relevant temperature of 303 K are summarized in Figure 1.5 (A). We compare those with the data on translational diffusion of ubiquitin, D_{tr}^{Ub}, shown in Figure 1.5 (B).

As seen from the plot, in the TIP4P-Ew simulations the self-diffusion coefficient of water is overestimated by 13.8 % (blue bar in Figure 1.5 (A). Similarly, the diffusion coefficient of Ub using this water model is overestimated by 16.7 % or 14.8 % (blue and light-blue bars in Figure 1.5 (B), corresponding to the NPT and NVE simulations, respectively).

Next, in the TIP4P-D simulations the self-diffusion coefficient of water is underestimated by 8.1 % (green bar in Figure 1.5 (A)). Similarly, the diffusion coefficient of Ub using this water model is underestimated by 7.4 % (green and light-green bars in Figure 1.5 (B), corresponding to the NPT and NVE simulations).

Finally, in the OPC simulations the self-diffusion coefficient of water is overestimated by 2.3 % (magenta bar in Figure 1.5 A). Similarly, the diffusion coefficient of Ub using this water model is overestimated by 4.3 % (magenta bar in Figure 1.5 (B), NPT simulations).

Table 2 — Summary of experimental and calculated diffusion coefficients for N-H4 and Ub. Direct scheme to calculate diffusion coefficients from the MD data is based on the $MSD(\tau)$ analyses and accounts for the box-size dependence of the results, see Figure 1.4. Four empirical schemes to predict the diffusion coefficients from a series of MD frames (HYDROPRO, Nygaard, Kirkwood-Riseman, and HullRadSAS schemes) are detailed in section 1.3. The experimental results for D_{tr}^{N-H4} and D_{tr}^{Ub} are from this study, while D_{rot}^{Ub} is from ref. [103], subject to minor adjustments related to temperature and solvent viscosity

Diffusion coefficient	Experimental value	Simulations														
		TIP4P-Ew					TIP4P-D					OPC				
		Direct	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS	Direct	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS	Direct	HYDROPRO	Nygaard	Kirkwood-Riseman	HullRadSAS
D_{tr}^{N-H4} (10^{-10} m^2/s)	1.71	2.14	1.75	1.72	2.34	2.05	1.59	1.56	1.62	1.96	1.8	1.81	1.63	1.66	2.11	1.9
D_{tr}^{Ub} (10^{-10} m^2/s)	1.62	1.89	1.61	1.66	–	1.75	1.5	1.6	1.65	–	1.74	1.69	1.61	1.66	–	1.75
D_{rot}^{Ub} (10^{7} rad^2/s)	4.87	5.50	4	–	–	4.14	4.35	3.87	–	–	4.05	4.94	3.98	–	–	4.13

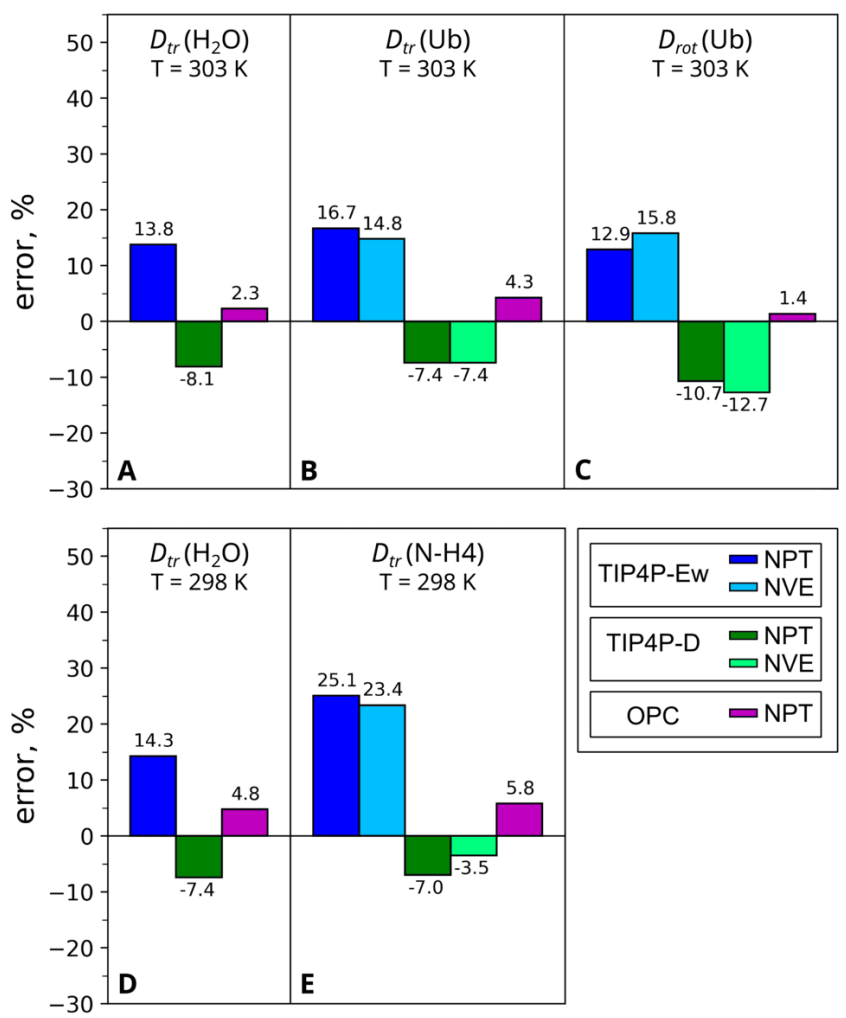


Figure 1.5 — The differences between the simulated and experimental values of diffusion coefficients for water, Ub and N-H4 (based on Table 1 and Table 2 and Table 6). The simulation and measurement conditions are annotated in the plot. Of note, the results for water self-diffusion coefficient at 298 K are very similar to those at 303 K (cf. panels A and D).

Thus, comparing the results inuresigs Fgiures 1.5 (A, B) we observe that they are essentially identical to within a couple of percentage points. This means that the simulated translational diffusion of Ub is indeed determined entirely by the viscosity of the MD solvent. In particular, when using the OPC solvent, which closely reproduces the experimental self-diffusion coefficient of water, the simulations also closely reproduce the experimental diffusion coefficient of Ub. In turn, this means that the MD-based scheme employed in our study can be used to successfully predict the translational diffusion of a globular protein — one just needs to use a water model with correct viscosity (such as OPC)

or, otherwise, apply a simple ex-post viscosity correction (recommendable for TIP4P-Ew and TIP4P-D).

This conclusion is supported by the rotational diffusion data for Ub, see Figure 1.5 (C). As can be seen from the plot, the predicted values of the rotational diffusion coefficient can be fully explained by the variations in solvent viscosity. In particular, when using the OPC solvent, the predicted diffusion coefficient $D^{Ub}_{(rot,0)}$ reproduces the experimental result to within 1.4 % (magenta bar in Figure 1.5 (C)). Otherwise, in the case of TIP4P-Ew or TIP4P-D solvent one needs to apply a simple viscosity correction in order to obtain a similarly accurate result.

In the above discussion, we observe that the bars in Figure 1.5 (A) are very similar in magnitude to the bars in Figures 1.5 (B. C). While some deviations are seen in the graphs, they are small and largely reflect the statistical uncertainty in the MD-based calculations (cf. green and light-green bars in Figure 1.5 (C), corresponding to the NPT and NVE simulations, respectively). Another source of uncertainy is the experimental error. As already indicated (see the previous section), the data entries in Figure 1.5 can be deemed consistent so long as they fall within several percentage points of each other.

Next, we turn to the N-H4 data, which are of prime interest to us in the context of this study. Since N-H4 in solution is disordered, we expect that the accuracy of the MD predictions for this peptide depends in this case on two factors: viscosity of the simulated solvent (already discussed for Ub) and the characteristics of the simulated conformational ensemble (see Figure 1.3).

The deviations between the MD-predicted and experimental diffusion coefficients for water and N-H4 at the relevant temperature of 298 K are graphed in Figures 1.5 (D, E). Comparing the data from water simulations with those from the N-H4 simulations, we immediately notice a big difference between the respective TIP4P-Ew results. While this water model overestimates the water self-diffusion coefficient by 14.3 % (blue bar in Figure 1.5 (D)), it overestimates the N-H4 diffusion coefficient by a much wider margin, 25.1 % or 23.4 % (blue and light-blue bars in Figure 1.5 (E)). The mismatch is too large to be explained away by the experimental error or statistical uncertainty of the MD simulations. Hence, we are led to conclude that the reason lies with the modeling of the peptide conformational ensemble in the N-H4 simulations.

Indeed, considering the results in Figure 1.3 we observe that the N-H4 peptide in the TIP4P-Ew solvent tends to adopt compact conformations. These compact conformational species diffuse more rapidly, resulting in the higher-than-expected translational diffusion coefficient. Thus, our analysis identifies a problem with the modelling of the N-H4 conformational ensemble in TIP4P-Ew water. Additional relaxation-based evidence that supports this interpretation is shown below.

Returning to the discussion of Figures 1.5 (D, E), we note for TIP4P-D and OPC models the data on self-diffusion of water are compatible with the data on diffusion of N-H4. Specifically, in the case of TIP4P-D model the self-diffusion coefficient of water is underestimated by 7.4 % (green bar in Figure 1.5 (D)), while the diffusion coefficient of N-H4 is underestimated by 7.0 % or 3.5 % (green and light-green bars in Figure 1.5 (E)). Finally, in the case of the OPC model the self-diffusion coefficient of water is overestimated by 4.8 % (magenta bar in Figure 1.5 (D)), while the diffusion coefficient of N-H4 is overestimated by 5.8 % (magenta bar in Figure 1.5 (E)).

The simple explanation is that both water models, TIP4P-D and OPC, are reasonably successful in modeling of the N-H4 conformational ensemble. Specifically, it can be suggested that the true $P(R_g)$ distribution for this peptide is somewhere in between the ones predicted by the TIP4P-D and OPC simulations (cf. Figure 1.3). In this situation, the MD-derived diffusion coefficient of N-H4 is determined mainly by solvent viscosity, which is what we observe for these two water models.

In order to corroborate these observations, we turn to the ^{15}N relaxation data [49]. These data include a set of per-residue longitudinal relaxation rates R_1 and transverse cross-correlated (dipolar-CSA) relaxation rates η measured at the temperature of 298 K. To draw a comparison, we have also calculated the R_1 and η rates using the MD trajectories reported in this study (see 1.3). Both experimental and calculated results are shown in Figure 1.6.

Of particular interest to us are the cross-correlated relaxation rates η Figure 1.6 (B). Unlike the longitudinal rates R_1, the transverse rates η have a simple dependence on motional time constants — namely, any slowing of protein dynamics leads to increases in η [104]. Let us first discuss the simulations employing TIP4P-Ew water. As already demonstrated, the viscosity of TIP4P-Ew water is substantially lower than the experimentally measured

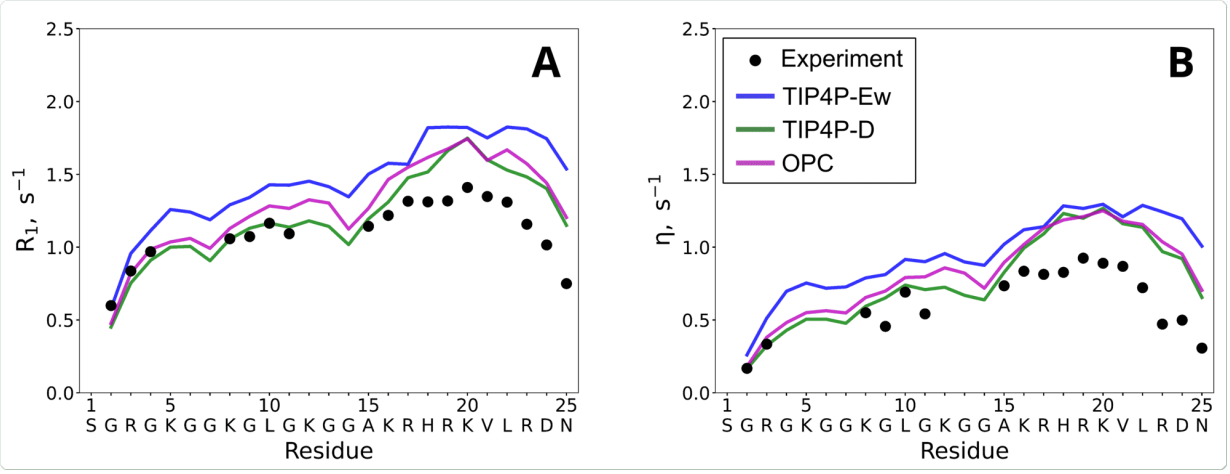


Figure 1.6 — Simulated and experimental ^{15}N relaxation rates (A) R_1 and (B) η in the N-H4 peptide. The notations are described in the figure legend. The simulated data are from the NPT trajectories in the small box. We have found that the calculated relaxation rates are essentially independent of the box size (see Figure 1.14, which is understandable since the spin relaxation in the N-H4 peptide is dictated by its extensive conformational dynamics rather than the overall tumbling

viscosity, cf. Table 1. In this situation, one should expect that the simulated dynamics of the N-H4 peptide in the TIP4P-Ew water should be faster than it is in reality. In turn, this implies that the simulated η rates should be lower than the experimentally measured ones. However, these expectations are not borne out by the actual results — in fact, the simulated η values are substantially *higher* than the experimental values (blue profile vs. the black circles in Figure 1.6 (B)).

Therefore, it remains to conclude that the results in Figure 1.6 (B) are influenced by the details of the N-H4 conformational ensemble. Recall that the ensemble observed in TIP4P-Ew simulations involves a significant proportion of loosely structured (hydrogen bonded) species and overall appears to be exceedingly compact (see subsection 1.3.2). The conformational dynamics of N-H4 is obviously slowed down in such transiently structured compact species, resulting in the elevated η rates. The effect is significant enough to overcompensate for the low viscosity of the TIP4P-Ew water, producing the simulated rates that are greater than the experimental values by a factor of ca. 1.5, see Figure 1.6 (B). Thus, the relaxation evidence appears to support our diffusion-based findings, indicating that the conformational ensemble of N-H4 peptide in the TIP4P-Ew water is exceedingly compact. This is also in line

with the previous knowledge that the classical water models tend to produce unrealistic «collapsed» models of disordered proteins [16].

Finally, returning to the discussion of Figure 1.6 we note that both TIP4P-D and OPC water produce a better description of the N-H4 relaxation rates than TIP4P-Ew. In fact, they accurately capture the rates in the more disordered N-terminal portion of the peptide, while still overestimating the rates in the C terminal segment (green and magenta profiles vs. the black circles in Figure 1.6). Given the smallness of the difference between the relaxation-rate predictions from the TIP4P-D and OPC simulations, it is hardly possible to favor one model over the other (i.e. it is impossible to tell which one provides a more realistic description of the N-H4 conformational ensemble, see Figure 1.3). While both diffusion and relaxation analyses point at the problem with the TIP4P-Ew water, further studies are needed to assess the relative performance of the TIP4P-D and OPC models.

1.2.5 Empirical methods to predict diffusion coefficients

In this chapter, we present the direct scheme to determine protein translational diffusion coefficient from the MD simulations. As a simple alternative, one can rely on the program HYDROPRO [70] which is intended to calculate hydrodynamic properties of macromolecules based on their coordinates. Of particular significance to us, in a number of studies this program has been used to predict the hydrodynamic radii and translational diffusion coefficients of disordered proteins. Specifically, HYDROPRO was applied to a set of MD-simulated IDP conformers and the results were subsequently averaged to obtain the predicted values of R_h and D_{tr} [82; 105—108]. The same approach has also been used for various ensemble models of IDPs generated by means other than MD simulations [109—111]. Furthermore, a few years ago, Nygaard et al. proposed a simple empirical parameterization to express the relationship between the average radius of gyration R_g and the HYDROPRO-predicted R_h for simulated or otherwise constructed conformational ensembles of IDPs [71]. This parametrization has also been used to predict the translational diffusion parameters of the MD-simulated IDPs [88; 89].

Strictly speaking, the approach whereby HYDROPRO is used to predict D_{tr} of disordered protein, lacks any solid theoretical foundation. The description of HYDROPRO clearly states that it is intended to calculate the hydrodynamic properties of rigid macromolecules [70]; hydrodynamic calculations on flexible macromolecules require different approaches [112; 113].

In effect, application of HYDROPRO to conformational ensembles of IDPs implies that all conformers are «frozen» and preserve their shape while diffusing in solution. This is clearly different from the real-life situation where the peptide's diffusion involves continuous conformational rearrangements. Intuitively, we anticipate that diffusion of the «frozen» conformers should be, on average, slower than that of the conformationally mobile peptide. Indeed, numeric simulations of simple polymer chains support this conjecture [114; 115].

To directly test this point, we have designed a special restrained simulation of the N-H4 peptide. Specifically, we regenerated a series of five hundred 10-ns trajectories representing N-H4 in OPC water in a medium-sized box. In each of these short trajectories, a large number of «soft» distance restraints have been imposed on the peptide so that it preserved its initial conformation during the simulation (see 1.3). The obtained set of short trajectories is thus representative of the diverse conformational ensemble of N-H4, but all of the simulated conformers are essentially rigid («frozen») and diffuse as such. As it turns out, this model leads to the D_{tr}^{N-H4} value which is 12 % lower than our original (unrestrained) result. The same kind of restrained simulation in a small-size box leads to the D_{tr}^{N-H4} value which is underestimated by 8 %. These findings directly confirm our notion that the ensemble consisting of «frozen» conformers shows slower diffusion rate compared to the fully dynamic peptide model.

Based on the above discussion, we expect that application of HYDROPRO to our original (unrestrained) MD data should lead to a similar underestimation of diffusion coefficients. To address this question, we have conducted the HYDROPRO calculations on 5000 frames from the small-box trajectories of N-H4 in the TIP4P-Ew, TIP4P-D and OPC water. The calculations used the experimental values for water viscosity and water density at 298 K, with other input parameters set to default values. The results were compared to the experimental D_{tr} value as measured in this work; the deviations between the

HYDROPRO-based predictions and the experiment are illustrated in Figure 1.7 (A).

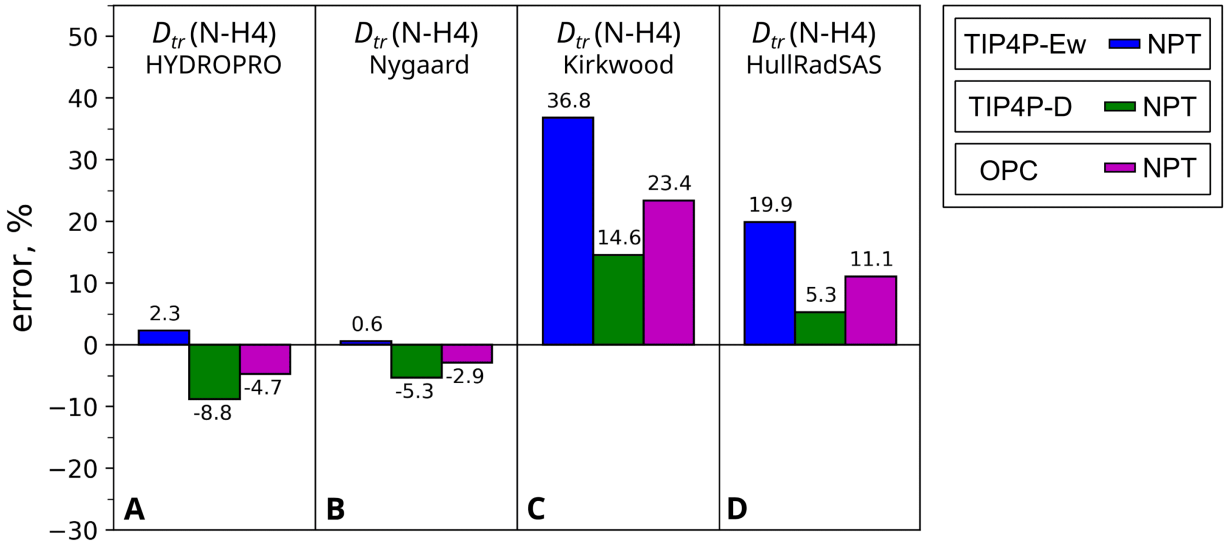


Figure 1.7 — The differences between the calculated and experimental values of diffusion coefficients for N-H4. The calculations use 5000 frames from the small-box NPT trajectories of N-H4 in TIP4P-Ew, TIP4-D and OPC water (blue, green and magenta bars, respectively). The results from medium- and large-box simulations are very similar to those shown in the plot, which is understandable given that the respective conformational ensembles are very similar, see Figure 1.12. The calculated D_{tr} values are from the following computational tools: (A) HYDROPRO program; (B) empirical Nygaard formula, which emulates the HYDROPRO results [71]; (C) Kirkwood-Riseman formula, formula (1.5); (D) HullRadSAS program. In the calculations B and C, the calculated values of R_h were converted into D_{tr} by means of the Einstein-Stokes formula (1.3). The alternative conversion method using the data from the reference molecule, formula (1.4), produces similar results (not shown).

Inspection of the HYDROPRO-based results in Figure 1.7 (A) suggests that TIP4P-Ew is the best water model, which faithfully reproduces the experimentally measured diffusion coefficient of N-H4. Indeed, the deviation from the target for this water model is only 2.3 %, compareable to the experimental uncertainty and appreciably better than that for the TIP4P-D and OPC models. This is in contrast to our direct analyses of the MD data (see section 1.3.2), which suggests that the use of TIP4P-Ew leads to the overly compact conformational ensemble for the N-H4 peptide; additionally, this water

model suffers from low viscosity. We believe that HYDROPRO results are in error, as determined by the following two factors. First, since HYDROPRO treatment tacitly assumes that the peptide molecules are rigid, it tends to underestimate the diffusion coefficients of the disordered species (see above). Second, at the same time, the simulations using TIP4P-Ew water produce an overly compact peptide ensemble, which results in HYDROPRO overestimating the relevant diffusion constants. The combination of these factors leads to error compensation, resulting in a false conclusion that TIP4P-Ew is the best choice of water model to simulate this disordered peptide. In this connection, one other factor should also be mentioned: HYDROPRO calculations rely on the user-supplied value of water viscosity (which is usually the experimentally measured value) and, therefore, are unaffected by the TIP4P-Ew viscosity per se.

From our perspective this latter aspect is both a strength and a weakness. On the one hand, the results are immune to the viscosity bias, which we have to deal with in our direct analyses of the peptide diffusion, see subsection 1.3.2. On the other hand, the HYDROPRO results fail to alert us to the problem with TIP4P-Ew viscosity (which can compromise other MD-based calculations, e.g. the calculations of spin relaxation rates).

To conclude the discussion of HYDROPRO, we note that this program shows very good accuracy in predicting D_{tr} for ubiquitin, see Figure 1.15 (A). The results are essentially independent of the water model used. Indeed, the structure of this small globular protein is only minimally sensitive to the type of the water model employed in the MD simulations (see section 1.3.2). At the same time, HYDROPRO performs rather poorly in predicting D_{rot} of ubiquitin, registering the errors of up to 20 %, see Figure 1.15 (D). This fact, which has also been noted by others [116], probably reflects the limitations of the hydrodynamics models used by HYDROPRO.

It should also be noted that the empirical relationship between R_g and the HYDROPRO-predicted R_h due to Nygaard et al. holds well for the proteins at hand, cf. Figure 1.7 (A) vs. Figure 1.7 (B) and Figure 1.15 (A) vs. 1.15 (B). This means, however, that our criticism of the approach whereby HYDROPRO is used to predict translational diffusion coefficients of IDPs also applies to the Nygaard's formula.

In conclusion, we recommend against using HYDROPRO on ensemble models of disordered proteins because this program is not designed for disordered proteins and can produce misleading results.

Very recently, Lindorff-Larsen and co-workers proposed two alternative schemes to predict translational diffusion parameters of difosrdered proteins based on their ensemble models. For a number of IDPs of different size, the authors have generated conformational ensembles by using either the well-known program Flexible-Meccano [117] or their original Langevin simulations using the coarse-grained force field CALVADOS [118]. The ensembles were subsequently reweighted using the experimental SAXS data. The authors then sought to establish the relationship between these ensembles and the R_h (D_{tr}) data from the PFG-NMR diffusion measurements.

To this end, the authors initially invoked a generalistic Kirkwood-Riseman model pertaining to hydrodynamics properties of flexible polymers [119]. As it appeared, this model was capable of accurately reproducing the experimental R_h values based on the ensemble models at hand, subject to some reservations concerning the protein size [72]. However, shortly thereafter the experimental data have been re-assessed and it was found that Kirkwood-Riseman formula actually tends to overestimate the translational diffusion coefficients of IDPs [73].

As an alternative, the same investigators proposed to use the recent program HullRadSAS [120]. When applied to the ensemble models at hand, this approach reproduces the experimental R_h values with a fairly good accuracy [73]. It should be noted, however, that HullRadSAS is a program for hydrodynamics calculations, which is conceptually similar to HYDROPRO. Just like HYDROPRO, it is not intended for disordered proteins. Therefore, the use of HullRadSAS in this context should be regarded as an empirical solution to draw a bridge between the givent set of models and the R_h (D_{tr}) data.

As a part of our study, we have tested both Kirkwood-Riseman formula and HullRadSAS scheme on our data (see section 1.3). The results from Kirkwood-Riseman interpretation appear to be unsatisfactory, see Figure 1.7 (C). In accordance with the latest observations [73], this method overestimates the diffusion coefficients of IDPs. While our rigorous analysis of the N-H4 diffusion in the MD simulations suggests that the trajectories recorded in TIP4P-D and OPC water achieve a near-quantitative accuracy in modeling the translational diffusion of this peptide (see section 1.3.2), the application of the

Kirkwood-Riseman formula to the MD snapshots leads to large errors of 14.6 % and 23.4 % for the two respective water models.

Turning to the very recent HullRadSAS scheme, see Figure 1.7 (D), we note that the results are indeed more in line with the direct MD-based analysis. Nevertheless, they still fall short of the quantitative agreement. Specifically, if we factor out the trivial viscosity effects, we expect that the conformational ensembles obtained from the TIP4P-D and OPC simulations should accurately reproduce the experimental diffusion characteristics of N-H4, see Figure 1.5 (E, D). Instead, the predictions by HullRadSAS overestimate the experimental D_{tr} value by 5.3 % and 11.1 % for TIP4P-D and OPC, respectively. This may lead one to incorrectly conclude that the OPC model is poorly suited for modeling of disordered proteins.

Given that HullRadSAS has not been designed to work with flexible IDPs, we repeat our call for caution which we have previously made with regard to HYDROPRO. If the goal is to validate an MD model of a disordered peptide, it is safer to directly extract the diffusion coefficient from the MD simulations (such as demonstrated in this work) rather than rely on empirical tools to predict D_{tr} based on a selection of MD snapshots.

1.2.6 Accelerating MD simulations of protein diffusion

As discussed above, direct determination of D_{tr} using the $MSD(\tau)$ metric is preferable to various empirical schemes that make predictions based on the coordinates of the MD-simulated conformers. However, the downside of the direct approach is that it is time-consuming. We have already shown that the computational time can be reduced several fold by recording fragmented trajectories of the peptide in the medium- and large-size water boxes. In this section we consider the possibility that further time savings can be achieved by using a longer integration timestep. To obtain some insight into this problem, we treat the simple test case, that of ubiquitin.

Toward this goal, we have recorded three additional series of Ub trajectories using longer integration timestep: $\Delta t = 2$ fs, $\Delta t = 4$ fs (with the HMR scheme applied to Ub, but not to water, as per the Amber manual

recommendation) and $\Delta t = 4$ fs (with the HMR scheme applied to both Ub and solvent water). Each series consists of the simulations in small-, medium- and large-size boxes and has a combined length of 6 μs. Aside from the timestep and the HMR scheme, the simulation protocol is identical to the one described in section 1.3. One of the two more successful solvent models, OPC, has been selected for these simulations. The results are summarized in Table 3.

The inspection of the data in Table 3 shows that making a transition from $\Delta t = 1$ fs to $\Delta t = 2$ fs causes moderate increases in the Ub diffusion coefficients, on the order of several percentage points. On the other hand, an attempt to use a longer timestep, $\Delta t = 4$ fs, in conjunction with the HMR scheme on Ub leads to more substantial increases, on the order of 15-20 %. At the same time, if the HMR scheme is applied to both Ub and water, the results from $\Delta t = 4$ fs simulations are brought back in line with the original 1 fs results.

This outcome is apparently at odds with the Amber default setup, where the HMR scheme is only applied to the protein molecule [121]. The rationale for this default setting is that the water molecules are already constrained to their rigid geometry via the SETTLE algorithm [122] and, therefore, do not need to be additionally constrained via the HMR. However, it has been previously pointed out that the SETTLE algorithm does not necessarily eliminate all high-frequency motions in water [123]. In fact, there are short-range nonbonded interactions between proximal water molecules, which can give rise to such high-frequency motions [123]. We argue that the application of the HMR scheme to both Ub and solvent helps to suppress these high-frequency motional modes and thus improve the accuracy of the MD simulations using the extended 4-fs timestep.

Our preliminary studies indicate that this conclusion also holds for pure water, as well as the intrinsically disordered N-H4 peptide. It appears that the simulations using $\Delta t = 4$ fs are feasible for these systems and produce the accurate diffusion parameters, but if and only if the HMR scheme is applied to the entire simulation cell. If confirmed, this result means that the direct MD-based procedure to quantify protein diffusion can be completed in one-quarter of the time that has been expended in this work.

Table 3 — Diffusion coefficients of Ub from the additional series of MD simulations recorded with different integration timesteps Δt and partial or full use of the HMR scheme. Other than these integration parameters, all simulations employed the same MD protocol as described in the section 1.3. The NPT thermodynamic ensemble and the OPC solvent model have been used. The extrapolation of the calculated diffusion coefficients to the limit of an infinitely large water box is illustrated in Figure 1.16

Protein	Diffusion coefficient	Experimental values	Simulations			
			Δt = 1 fs no HMR	Δt = 2 fs no HMR	Δt = 4 fs HMR (protein)	Δt = 4 fs HMR (protein + H2O)
Ub	10^{-10} m^2/s	1.62	1.69	1.8	1.93	1.68
	10^7 rad^2/s	4.87	4.94	4.97	5.91	4.72

1.3 Materials and methods

1.3.1 MD simulations

The initial coordinates of N-H4 peptide (amino-acid sequence SGRGKGGKGLG KGGAKRHRKVLRDN) were generated as described previously [49]. Briefly, we have built 2000 random N-H4 conformations using the server `http://unfolded.uchicago.edu` [124] and the program Scwrl4 [125]. All conformations were energy-minimized and then ordered according to their energies in GBneck2 solvent [126]. One structure was chosen at random from the central portion of the energy histogram and subsequently used in all simulations.

The initial structure was protonated in accordance with the experimental pH 4.0 using the program PROPKA [127]. The low pH was originally chosen to minimize amide proton exchange with solvent; the protonation of the peptide at pH 4.0 is the same as at physiologically relevant pH 7.2 except for a single histidine residue [49]. The simulations were conducted in Amber 20 (ff14SB) using three different water models: TIP4P-Ew, TIP4P-D and OPC. For each choice of water model, we have recorded the trajectories in solvation boxes of different size. Here we describe the simulations using smaller boxes; the procedure to record trajectories in medium- and large-sized boxes is explained later.

The truncated octahedral boxes were built using *SolvateOct* command in LEaP; the distance parameter (nominally, a minimal separation between the peptide and the boundary of the cell) has been set to 12 Å for TIP4P-Ew simulations or, otherwise, 15 Å for TIP4P-D and OPC simulations. The system was neutralized by adding nine Cl^- ions [128] and then energy-minimized with harmonic restraints applied to N-H4 heavy atoms. After that the system was brought to the target temperature of 298 K and equilibrated for 1 ns in NVE ensemble prior to the production run.

As already pointed out, in our simulations we opted for velocity rescaling thermostat, also known as Bussi thermostat [63], which preserves the native-like dynamics of the system. For this purpose, we have implemented Bussi thermostat as a part of the official Amber 20 release (option $ntt = 11$). The pressure was

controlled by Berendsen barostat [129] with the coupling time of 1 ps. The nonbonded interactions were calculated with a cutoff of 11 Å. The particle mesh Ewald summation scheme has been employed to treat long-range electrostatic interactions with the default parameters for grid spacing and spline interpolation. Bonds involving hydrogens have been restrained by means of the SHAKE algorithm [130]. The integration time step was 1 fs (motivated by a separate series of NVE simulations, see below). The coordinates were stored every 1 ps. The simulations were conducted using in-house GPU workstations equipped with NVIDIA GeForce GTX 1080, RTX 2080 Ti and RTX 3080 cards. The length of the small-box N-H4 simulations employing three different water models was 5 µs apiece.

Note that the above description refers to the MD simulation of the disordered peptide in a periodic boundary cell of a limited size. Does this have any constraining effect on the N-H4 exploring its conformational phase space? The answer depends on a number of encounters between the peptide and its periodic images during the course of the simulations. To address this issue, we have analyzed the trajectories and calculated the fraction of all frames featuring close encounters between the master copy of N-H4 and its periodic images (where at least a pair of atoms come to within 5 Å of each other). As it turns out, the proportion of such frames is only a fraction of percentage point, viz. 0.06 %, 0.36 % and 0.17 % for trajectories in TIP4P-Ew, TIP4P-D and OPC water, respectively. Hence, we conclude that the crowding effect in our simulations must be minimal and can be safely ignored.

As already pointed out, the procedure to determine D_{tr} and D_{rot} relies on several MD trajectories recorded in water boxes of increasing size. While the simulations using relatively small solvent boxes (see above) are reasonably fast, their counterparts involving medium- and large-sized boxes are far more time consuming. As a time-saving device, we propose a special scheme to record these latter trajectories, see Figure 1.8. In brief, we extract the N-H4 coordinates from the frames at time 0, 10, 20, ..., 4990 ns of the small-box trajectory and then place these conformers into bigger boxes. Specifically, medium-sized boxes were generated by adding, via SolvateOct, a water shell with the minimal thickness of 24 Å, whereas the large boxes were generated with the water layer of 48 Å. The so prepared simulation cells were then equilibrated as described above and used to start short trajectories with the length of x ns (see Figure 1.8).

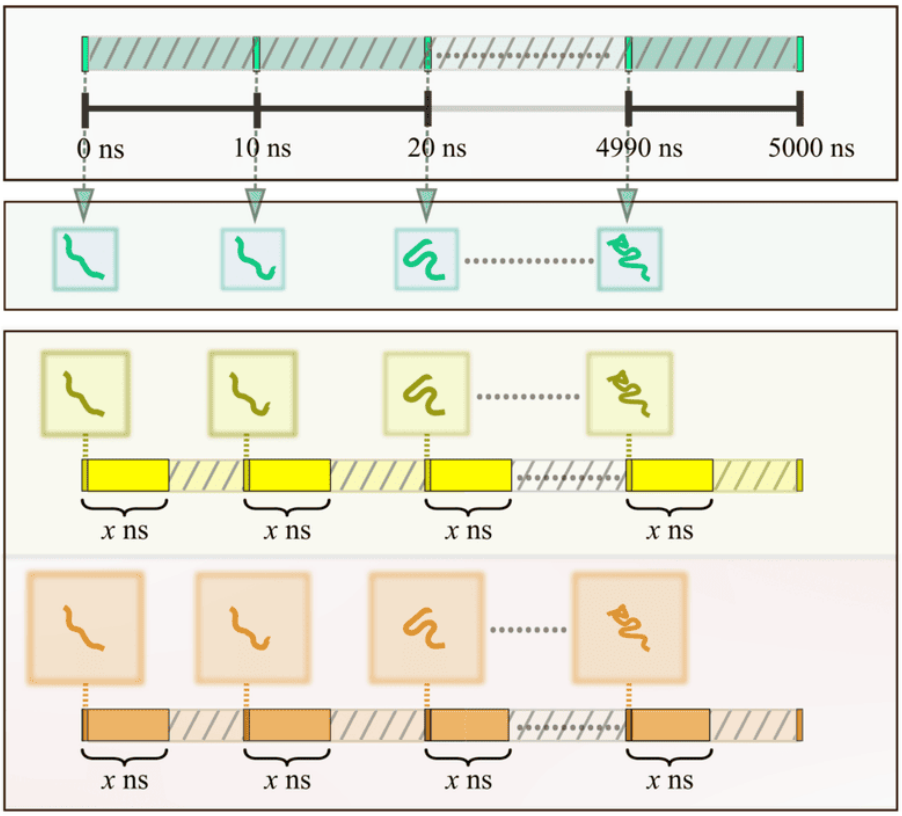


Figure 1.8 — Schematic design of N-H4 simulations using small-, medium- and large-size water boxes. The short trajectories were initially recorded with the length of $x = 5$ ns and later extended to $x = 10$ ns

In this work, we recorded a series of trajectories with $x = 5$ ns. This scheme offers substantial savings in computational time. First, there is a two-fold gain due to the reduced overall length of the simulations, see Figure 1.8. Second, there is a several-fold gain from using a network of GPU computers equipped with a queuing system (in our case, SLURM), which allows one to quickly record a series of short trajectories. At the same time, the described scheme ensures satisfactory conformational sampling for the medium- and large-box simulations, as inherited from the continuous 5 µs trajectory of N-H4 in a smaller water box. In order to demonstrate the results are converged, we have later extended $x = 5$ ns series of simulations to $x = 10$ ns.

Alongside with NPT simulations employing Bussi thermostat, we have also recorded a series of NVE simulations of N-H4 peptide in TIP4P-Ew and TIP4P-D water. It is well-known that NVE simulations represent «true» dynamics of the system, free from potential interference by thermostating algorithms. At the same time, NVE simulations often develop a substantial temperature drift, which makes them ill-suited for routine applications [131;

132]. In the context of our work, NVE simulations have been conducted as a control, with the aim to validate the NPT results using Bussi thermostat. The protocol to record NVE trajectories was identical to the one described above, including the integration timestep of 1 fs (intended to reduce the temperature drift). Using the scheme shown in Figure 1.8, we only need to worry about the temperature drift during the 5-μs NVE simulations in the smaller box (since the simulations in medium- and large-sized boxes are recorded in short segments of $x = 5$ or 10 ns and hence do not suffer from the temperature drift). As it happens, the observed temperature drift in 5 μs NVE simulations of N-H4 using small water boxes is modest: on average, the temperature is increased by 1.2 K in the TIP4P-Ew trajectory and by 0.5 K in the TIP4P-D trajectory. Although these effects are not inconsequential, they are tolerable for the purpose of our analyses, which is to validate the results of the NPT simulations.

The summary of all N-H4 trajectories recorded in this study is given in Table 4. The net length of these trajectories amounts to 75 μs.

Separately, in order to investigate the effect of N-H4 internal mobility on its diffusion properties (in relation to HYDROPRO-based predictions of D_{tr}) we have conducted a special restrained simulation of N-H4. Specifically, we regenerated the set of five hundred 10-ns trajectories of N-H4 in OPC water in the medium-sized box. Each of these trajectories was recorded with soft restraints imposed on all pairwise interatomic distances (the total of 73536 restrains). The restraint potential was a well with parabolic bottom, $k(d - d_0)^2$ (extending from 0.99 d_0 to 1.01 d_0, where d_0 is the interatomic distance found in the initial frame, with k set to 1 kcal/mol $\cdot$ Å^2) and linear sides beyond that [66]. The combined effect of all restraints was to maintain the peptide conformation close to its initial conformation (which is specific to each individual 10-ns simulation); the root mean square deviation of the peptide coordinates relative to the initial frame did not exceed 0.15 Å. The temperature, pressure and volume of the restrained simulations remained on target and stable; the net restraint energy was a small fraction, 0.11 %, of the total energy. The same kind of simulations were also conducted in a small box; in doing so, the continuous small-box trajectory of N-H4 in OPC water was divided into 10-ns segments and used to seed five hundred 10-ns restrained simulations.

In addition to N-H4, we have also recorded a series of trajectories for the popular model protein ubiquitin, which has been chosen in our study as a

control system. The initial coordinates were from the structure 1UBQ [133] protonated using tleap (according to PROPKA, the protonation pattern of Ub remains the same between pH 4.5 and 6.5). The simulation temperature was 303 K; other MD parameters were the same as in the N-H4 simulation. The design of the simulations was similar to the one described above, see Figure 1.8, i.e. the trajectory in a small box was recorded in one piece, whereas the trajectories in medium and large boxes were recorded as collections of multiple short trajectories with $x = 5$ ns (subsequently extended to 10 ns). Since the requirements regarding conformational sampling in Ub simulations are rather minimal, we have reduced the duration of the simulations from 5 to 2 µs. The same tactics was used for NVE simulations of ubiquitin in TIP4P-Ew and TIP4P-D water. The temperature drift in the small-box NVE trajectories of Ub proved to be insignificant (0.4 and 0.2 K, respectively), thus facilitating the type of the procedure shown in Figure 1.8. The summary of Ub simulations can be found in Table 5; their net length amounts to 30 µs.

To obtain a handle on water viscosity for TIP4P-Ew, TIP4P-D and OPC models, we also recorded a set of trajectories for pure water in the simulation boxes of increasing size. The dimensions of the small, medium and large boxes were chosen to be the same as in Ub simulations. The length of each trajectory was 0.15 µs; the net length of all water simulations was 1.35 µs.

Additionally, to explore the significance of the integration timestep, we simulated Ub in the OPC water using 2 fs and 4 fs timestep with or without the hydrogen mass repartition scheme (HMR) [134].

1.3.2 Processing of MD data

As a first step, the MD trajectories using periodic boundary conditions were unwrapped, paying special attention to box-volume fluctuations in the NPT simulations [135]. To this end we used the in-house script based on the python library pyxmolpp2.

The translational diffusion coefficients D_{tr} were calculated in a standard manner using mean square displacement (MSD) of the protein's center of

mass [136]. For collection of N trajectories with the individual length of x ns, the MSD was computed by formula

$$MSD(\tau) = \frac{1}{N \cdot n} \sum_{i=1}^{N} \sum_{j=1}^{n} |\vec{r}_i(t_j + \tau) - \vec{r}_i(t_j)|^2 \tag{1.1}$$

where $\overrightarrow{r_i}$ defines the protein's center of mass in the i-th trajectory, t_j and $t_j + \tau$ are a pair of time points separated by the interval τ, and n is a number of such pairs within the trajectory of length x. Whenever a single long trajectory is used instead of the collection of short trajectories, the formula in formula (1.1) is reduced to the familiar simple expression [136].

The obtained dependencies $MSD(\tau)$ have been fitted using a linear fitting function to extract D_{tr}. In doing so, it is important to choose the appropriate τ interval over which the fitting is performed. This interval should not extend to large τ values where the accuracy of the $MSD(\tau)$ curve suffers from increasingly poor statistics that manifests itself in large correlated errors [137; 138]. On the other hand, one can argue that small τ values should also be left out. In this connection, the role of internal protein dynamics is particularly relevant (discussed below).

Conformational transitions in a disordered protein lead to a (limited) displacement of protein's center of mass. This effect can be thought of as restrained translational diffusion. The characteristic time constants of this diffusion process can be gleaned from N-H4 simulations, where the two dominant motional modes have been observed: the fast mode on ca. 100 ps time scale and the slow mode on ca. 1 ns time scale [49]. Therefore, one can expect that conformational dynamics of N-H4 makes a certain contribution to the simulated dependence $MSD(\tau)$ in the sub-nanosecond time window

Indeed, some evidence of this behavior can be seen in our MD data (an example is shown in the inset of Figure 1.9 (E)). On the other hand, we note that the experimentally measured D_{tr} values correspond to very long τ intervals, on the order of tens or hundreds of milliseconds, and thus must be completely insensitive to protein internal motions. Therefore, our goal is to fit $MSD(\tau)$ such as to leave out the presumed small contributions from fast internal dynamics.

Considering these criteria, we have chosen the fitting interval [100 ps, 1 ns] to fit $MSD(\tau)$ profiles and determine D_{tr} (illustrated in Figure 1.9). We have also tested other choices of the fitting interval and found only a small

amount of random variation in the resulting D_{tr} values, thus indicating that the scheme is sufficiently robust (Figure 1.9). We have also tested the convergence of our scheme by comparing the D_{tr} values from $x = 5$ ns and $x = 10$ ns sets of trajectories. The so obtained D_{tr} values are in good agreement with each other, confirming that the results are converged (see Figure 1.10 (A)). In what follows, we use the $x = 10$ ns data to determine D_{tr} with maximum possible accuracy. Likewise, the more statistically sound $x = 10$ ns data have also been used for all other calculations in this chapter.

The procedure to calculate D_{tr} for ubiquitin is the same as described above for N-H4. The procedure to calculate D_{tr} for water has a similar design as well. In water molecules, internal dynamics occurs on a femtosecond time scale; therefore, we have used the short-time fitting interval [1 ps, 100 ps] such as to reduce the amount of statistical uncertainty affecting D_{tr} [137; 138]. The D_{tr} values have been determined for 100 water molecules chosen at random from within the water box and subsequently averaged.

Ultimately, to predict the translational diffusion coefficient, we plot the obtained D_{tr} values as a function of inverse simulation box size. Any linear size of the truncated octahedron can be used for this purpose, e.g. the edge length a; we have chosen the diameter of the inscribed sphere, $L = \sqrt[6]{a}$, as a characteristic size of the box. Conveniently, the value of L is listed in the last line of the rst7 coordinate file.

Of note, N-H4 simulations employing medium and large cells involve the boxes of somewhat variable size (depending on the starting peptide conformation, see above). For example, considering large boxes in the five hundred short OPC trajectories, we find $L = 142 \pm 9$ Å. Of interest to us, the inverse size, $1/L$, varies according to 0.0071 ± 0.0004 Å^{-1}. When extrapolating to an infinitely large box, the average $1/L$ value has been used, 0.0071 Å^{-1} (see Figure 1.4). In principle, it would be straightforward to use the boxes of unified size (e.g. corresponding to the largest of all large boxes generated in our treatment).

Comparing D_{tr} values derived from the independently performed NPT and NVE simulations provides a good handle on MD-related uncertainty. In addition, we have also employed a jackknife method to estimate the uncertainties of MD-predicted diffusion coefficients. In doing so, the trajectories in small-sized boxes were divided into 10-ns segments, similar to trajectories in medium- and

large-sized boxes. Then for each box size we have discarded at random 20 % of all 10-ns segments and reprocessed the resulting redacted dataset to determine the translational diffusion coefficient. This procedure was repeated 1000 times; the standard deviation of the obtained D_{tr} distribution was taken to be the uncertainty of the D_{tr} prediction (reported in Table 6).

While D_{tr} has been determined for both N-H4 and Ub, the rotational diffusion coefficient D_{rot} is well defined only for Ub. To extract this parameter from the MD data, we use the following multi-step procedure. First, superpose (via the secondary-structure Cα atoms) the Ub molecules from all MD frames onto the Ub molecule from the first frame. Second, parameterize the above superposition operations via rotation matrices. Third, construct a pseudo-molecule containing 64 vectors (emulating N-H^N bonds) with near-uniform distribution on a unit sphere [139]. Fourth, apply the above rotation matrices to this pseudo-molecule (thus generating a pseudo-trajectory that encodes the protein's tumbling motion). Fifth, evaluate the time-correlation functions $g(\tau)$ for all 64 vectors by using formula

$$g(\tau) = \frac{1}{N \cdot n} \sum_{i=1}^{N} \sum_{j=1}^{n} P_2(\cos \chi_i(t_j, t_j + \tau)) \tag{1.2}$$

where $P_2(x)$ is the second-order Legendre polynomial, $\chi_i(t_j,t_j + \tau)$ is the angle between the vector's orientations at points in time t_j and $t_{j+\tau}$ in the i-th trajectory, and the averaging is over n pairs of time points separated by the interval τ and over N trajectories of x ns length [140]. To speed up the calculations, the inner sum in formula (1.2) has been evaluated using the fast Fourier transformation [141].

Sixth, combine the correlation functions from all 64 vectors, including the prescribed integration weights [139], to obtain $G(\tau)$. Seventh, fit $G(\tau)$ to a mono-exponential function to determine the tumbling time τ_{rot} and subsequently calculate $D_{rot} = 1((6\tau_{rot}))$. Similar to the treatment of translational diffusion, it is important to carefully select the fitting interval. We have investigated this aspect in some depth using long MD and BD (Brownian dynamics) simulations of ubiquitin (not shown) and concluded that the most accurate results are obtained when using the fitting interval [0, 2 ns]. The examples of $G(\tau)$ curves along with their best fits are illustrated in Figure 1.11.

The above procedure allows us to recover D_{rot} that is comparable to the one extracted from NMR relaxation studies. There are also other schemes that can be used to extract D_{rot} from molecular dynamics trajectories. For example, one can use MD data to calculate ^{15}N relaxation rates (cf. discussion below), analyze these simulated data to determine the anisotropic rotational diffusion tensor of ubiquitin [142], and from there make a transition to the isotropic diffusion coefficient D_{rot}. However, here we favor a simpler and more general approach [140], as described above.

Similar to translation, D_{rot} data also need to be extrapolated to the infinite-size box. In this case, the dependence of D_{rot} on the inverse box volume, rather than the linear size, is used to perform linear extrapolation [143]. In lieu of the volume one can use any quantity that is proportional to the volume; for the sake of convenience we have used L^3. The convergence of this procedure is illustrated in Figure 1.10 (B). The uncertainty of D_{rot} predictions has been estimated using the same jackknife procedure as described above for D_{tr}.

The scripts written in-house to extract D_{tr} and D_{rot} from an MD trajectory are available for download at `https://github.com/bionmr-spbu-projects/2023-UBQ-NH4-DIFFUSION`. This repository also contains installation and usage notes, as well as a short sample trajectory of Ub that can be used for the purpose of testing. The requisite python library pyxmolpp2, which has been developed in-house as a toolkit for processing of MD data, can be downloaded from `https://github.com/bionmr-spbu/pyxmolpp2`.

While the concept of rotational diffusion is not well defined for a disordered protein, a comparable dynamics information can be obtained from heteronuclear relaxation data. Here we used the MD simulations of N-H4 to calculate residue-specific ^{15}N longitudinal relaxation rates R_1 and CSA-dipolar cross-correlated cross-relaxation rates η [49]. The corresponding correlation functions were calculated using formula (1.2); in the case of cross-correlations, the angle $\chi$was taken to be the angle between the dipolar (N-H$^{\mathrm{N}}$) vector at time t and the unique axis of the axially symmetric ^{15}N CSA tensor at time $t + \tau$. When dealing with fragmented trajectories, the calculations were conducted on $x = 10$ ns dataset (same as for D_{tr} and D_{rot}).

The MD-derived correlation functions were fitted using the 4-exponential ansatz, $\sum_{k=1}^{4} c_k \exp(-\tau/\tau_k)$. The four characteristic times $_k$ have been constrained to the intervals [1 ps, 10 ps], [10 ps, 100 ps], [100 ps, 1 ns]

and [1 ns, 10 ns], while the weights c_k have been normalized to ensure that they sum up to 1.0. The fitting was performed using Levenberg-Marquardt algorithm implemented in SciPy function curve_fit [144]. The fitted $_k$ and c_k values were translated into spectral densities, which were in turn used to calculate R_1 and η [145]. In these calculations, we used the standard values for N-H^N bond length (1.02 Å), ^{15}N chemical shift anisotropy (-170 ppm), and the angle that the unique axis of the nitrogen CSA tensor makes with N-H^N bond (20∘) [146].

The HYDROPRO calculations were conducted on the individual frames from the small-box MD simulations (stride 1 ns) using the experimental water viscosity and density at 298 K or 303 K (for N-H4 and Ub, respectively) and other input parameters set to default values. The calculations using HullRadSAS [120] were performed in the same manner using the experimental water viscosity and otherwise in the default mode. The empirical relationship between the MD-derived average radius of gyration R_g and the HYDROPRO-calculated hydrodynamic radius R_h was used as given in the original report [71]. The R_h was converted into D_{tr} according to the authors' prescriptions via the Einstein-Stokes formula

$$D_{tr} = \frac{k_b T}{6\pi\eta_w R_h} \tag{1.3}$$

where k_b is the Boltzmann constant, T is the temperature, and eta_w is the water viscosity, or otherwise via the diffusion data from the reference molecule, 1,4-dioxane using formula

$$D_{tr} = D_{tr}^{dioxane} \frac{R_h^{dioxane}}{R_h} \tag{1.4}$$

where the hydrodynamic radius of dioxane $R_h^{dioxane}$ was assumed to be 2.27 Å [73], and the diffusion coefficient of dioxane in water $D_{tr}^{dioxane}$ at 298 K was taken to be 1.10 × 10^{-9} m^2/s [147]. The Kirkwood-Riseman formula was used in the following form [148]

$$R_h = 1\langle \frac{1}{N^2} \sum_{i \neq j} \frac{1}{r_{ij}} \rangle \tag{1.5}$$

where r_{ij} is the distance between the Cα atoms from i-th and j-th residues, N is the number of residues in the disordered protein chain, and the angular

brackets denote averaging over multiple MD frames. The transition from R_h to D_{tr} has been made using either formula (1.3) or formula (1.4).

1.3.3 NMR measurements

N-H4 peptide was synthesized by Pepmic Co. Ltd (Suzhou, China). The sample with peptide concentration 1 mM was prepared in 20 mM NaAc-d3 buffer containing 5 % D2O (pH 4.0). Recombinant Ub was expressed and purified as described elsewhere [149]. The sample was prepared with low protein concentration, 0.23 mM, to reduce the proportion of Ub dimers [150]; lyophilized Ub was dissolved in 20 mM NaAc-d3 buffer containing 5 % D2O (pH 6.0), approximating the conditions of ubiquitin dimerization study [150]. For the purpose of diffusion measurements, both N-H4 and Ub solutions were placed in D2O-susceptibility-matched Shigemi tubes with the sample volume 250 μL [151].

Translational diffusion measurements were conducted on Bruker Avance III 500 MHz spectrometer equipped with 5-mm BBI probehead with z-axis gradient. The experiments were carried out at 298 K for N-H4 and 303 K for Ub. The convection compensated double-stimulated echo sequence [152] with 3-9-19 WATERGATE [153] water suppression has been employed; the sequence code is based on Bruker pulse programs dstebpgp3s and stebpgp1s19. The standard Bruker smoothed square shape SMSQ10.100 with shape factor of 0.9 has been used for all gradient pulses. The duration of each component of the bipolar encoding/decoding gradient pulses was equal to 2.7 ms. Twenty spectra were acquired with the gradient amplitude ranging from 2.4 to 50 G/cm. The diffusion delay Δ was 0.1 s, the acquisition time was 3.6 s and the recycling delay was 5 s.

The aliphatic high-field proton signals (several spectral lines in the range from 0.93 to 0.73 ppm for N-H4 and a single line at -0.17 ppm for Ub) that are far removed from the residual water resonance have been chosen for diffusion coefficient determination. The integral intensity of these signals have been evaluated using the new algorithm developed in our laboratory [154], which is now available through the web server DDfit (Diffusion Data fit), `https://ddfit.bio-nmr.spbu.ru`. To extract the D_{tr} coefficient, the data were fitted using Stejskal-Tanner[155] with Jerschow-Müller modifications [152].

Prior to comparison with the MD data, the experimental results were corrected for a small fraction of D2O and sodium acetate in the buffer (assuming that D_{tr} is inversely proportional to solvent viscosity, cf. Einstein-Stokes. Furthermore, for ubiquitin the result was corrected for the presence of ubiquitin dimers [150] (which are responsible for 8 % of the NMR signal from the dilute sample at hand). Of note, this correction amounts to only 1.7 % of the measured D_{tr} value; in this sense, the experiment is rather insensitive to the modest proportion of dimers in the sample. The corrected D_{tr} values, corresponding to a monomeric protein in pure water, allow for direct comparison with the results of MD simulations.

The D_{rot} value for Ub is from the paper by Charlier et al., where a unique set of ^{15}N relaxation data at multiple magnetic fields has been collected for samples with different protein concentrations, including a highly dilute sample [103]. Under the conditions of their study, pH 4.5, the authors did not find any evidence of Ub dimerization, but observed nonspecific self-association behavior at high protein concentration. To avoid this effect, we have used the reported D_{rot} value measured at low protein concentration of 0.2 mM. This result was corrected for temperature (296.6 K in the experimental measurements), as well as the presence of D2O and acetate in the solvent, using the Einstein-Stokes The ^{15}N relaxation data for N-H4 peptide were used as originally reported [49].

1.4 Concluding remarks

Nowadays long peptides or smaller IDPs can be adequately modeled by means of conventional MD simulation techniques. It is therefore highly desirable to test and experimentally validate different MD models of such disordered systems. In particular, it would be useful to identify a model system, which can be investigated through concerted efforts of many research groups. For instance, peptides such as RS peptide [43; 72; 82; 156] or N-H4 [49; 99] investigated in this work can be used toward this goal. Such a thoroughly characterized disordered system could play a role similar to the one played by ubiquitin or lysozyme in strucutural and dynamic studies of globular proteins.

Measurements of translational diffusion by PFG-NMR can potentially provide a valuable piece of experimental data to test and validate MD models of disordered proteins. However, this type of analysis faces a number of hurdles. As it happens, D_{tr} is not a very sensitive parameter. This becomes clear when you consider the fact that D_{tr} of monomeric and homodimeric proteins differ by only a factor of $2^{1/3} \approx 1.26$. This lack of sensitivity makes it difficult to register small changes in compactness/extendedness of the IDP's conformational ensemble via the diffusion measurements.

Despite their wide popularity, PFG-NMR experiments are technically challenging, suffering from convection artefacts and baseline distortions caused by residual water signal [157]. On the computational front, prediction of D_{tr} from the MD data also presents a number of challenges. As shown in this work, it requires a judicious choice of the thermostating algorithm and multiple long MD simulations, including those in extra-large water boxes. To interpret the results, it is also necessary to accurately know the viscosity of the MD-simulated water for the water models at hand.

In addressing these problems, we have implemented the Bussi-Parrinello velocity rescaling thermostat in the MD simulation program Amber. We have also implemented a fragmentation scheme, which allows one to efficiently record MD trajectories in large-sized boxes while maintaining a good sampling of the peptide's conformational space. Our preliminary findings suggest that the simulations can be further accelerated by switching to the 4-fs integration step, with the proviso that the HMR scheme should be applied not only to the peptide but also to solvent. Separately, we have investigated the viscosity of TIP4P-Ew, TIP4P-D and OPC water and validated the results by modeling the diffusion of a small globular protein, ubiquitin.

The central theme of this work is a careful analysis of translational diffusion of the N-H4 peptide based on a series of NPT and NVE simulations with the net length of 75 μs. The analysis suggests that the simulations using the classical TIP4P-Ew water model produce an overly compact conformational ensemble for this disordered peptide. This conclusion is convincingly supported by the comparison of the simulated and experimental ^{15}N relaxation rates. Indeed, in the TIP4P-Ew trajectory the peptide is observed to form various «mini fold» arrangements, which are held together by opportunistic hydrogen bonds and

gradually morph from one shape to another, but remain recognizable for up to several hundreds of nanoseconds.

On the other hand, we have found that both TIP4P-D and OPC water models lead to the N-H4 conformational ensembles that are consistent with our experimental D_{tr} result. While the two ensembles are somewhat different, as characterized by the average R_g values of 15.3 and 13.8 Å, the diffusion analyses are not sufficiently sensitive to discriminate between them. Likewise, the ^{15}N relaxation rates calculated from the TIP4P-D and OPC trajectories are similar to each other and both show good agreement with the experimental results, such that we cannot prefer one water model over the other. Further data are needed to shed additional light on this problem.

As an alternative to the rigorous procedure to extract D_{tr} from the MD simulation data, various simplified methods have been widely used in this area. In particular, D_{tr} is often calculated by applying the program HYDROPRO to the MD-simulated conformers of the disordered protein. Here we have shown that this approach can lead to misleading results. For example, it identifies TIP4P-Ew as the water model that is best-suited to simulate the disordered N-H4 peptide, which is contrary to our (rigorous) findings. The main issue with the use of HYDROPRO is that it is designed to predict the diffusion properties of rigid biomolecules rather than the intrinsically flexible IDPs.

Very recently a number of empirical tools have been developed to predict the translational diffusion coefficients of IDPs based on their simulated conformational ensembles. These tools were calibrated on the ensembles informed by the experimental SAXS data and the experimental D_{tr} results from PFG-NMR experiments. We have found that these predictions can also be inaccurate, likely due to subtle experimental biases e.g. related to the use of 1,4-dioxane as a reference molecule [73]. In this sense, the first-principle predictions of D_{tr} from MD simulations of a disordered protein, such as demonstrated in this chapter, provide an important benchmark and validation point for future efforts in this area.

1.5 Supporting information

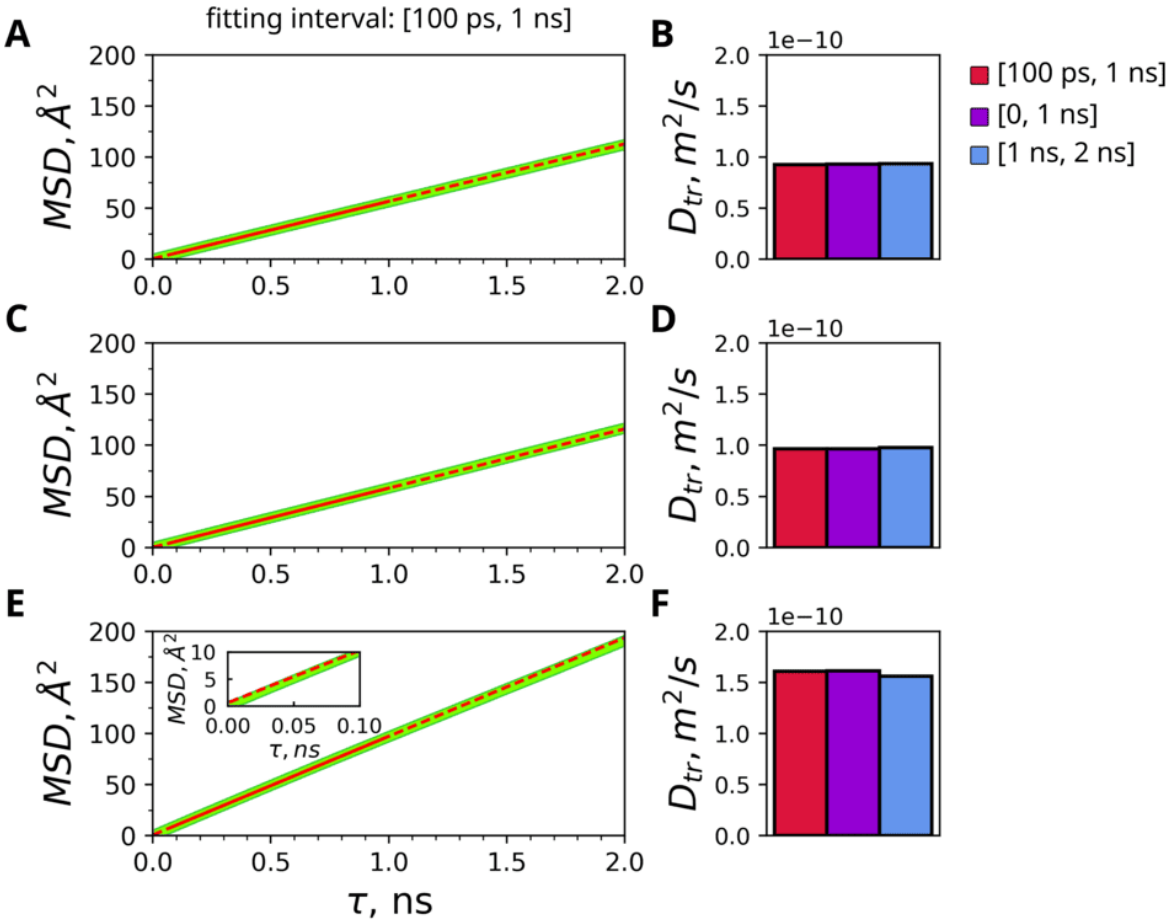


Figure 1.9 — Extracting D_{tr} values from MD simulations of the N-H4 peptide. The $MSD(\tau)$ curves obtained according to formula (1.1) (broad green lines) have been fitted using a linear fitting function $D_{tr} \cdot (\tau + b)$. The best fit is shown as a solid red line over the τ interval where the fitting has been performed; it is shown with a dashed line outside this interval. Displayed in the left column are the examples of the fits using the fitting interval [100 ps, 1 ns]. Shown in the right column are the summaries of the D_{tr} values obtained using different fitting intervals (color-coded as explained in the legend). (A, B) trajectory in TIP4P-D water, medium box, NPT ensemble; (C, D) trajectory in TIP4P-D water, medium box, NVE ensemble; (E, F) trajectory in TIP4P-Ew water, large box, NPT ensemble. The inset in panel E shows the initial portion of the $MSD(\tau)$ curve along with the best-fit line. Included in this figure are the simulations showing the most significant variations between the extracted D_{tr} values.

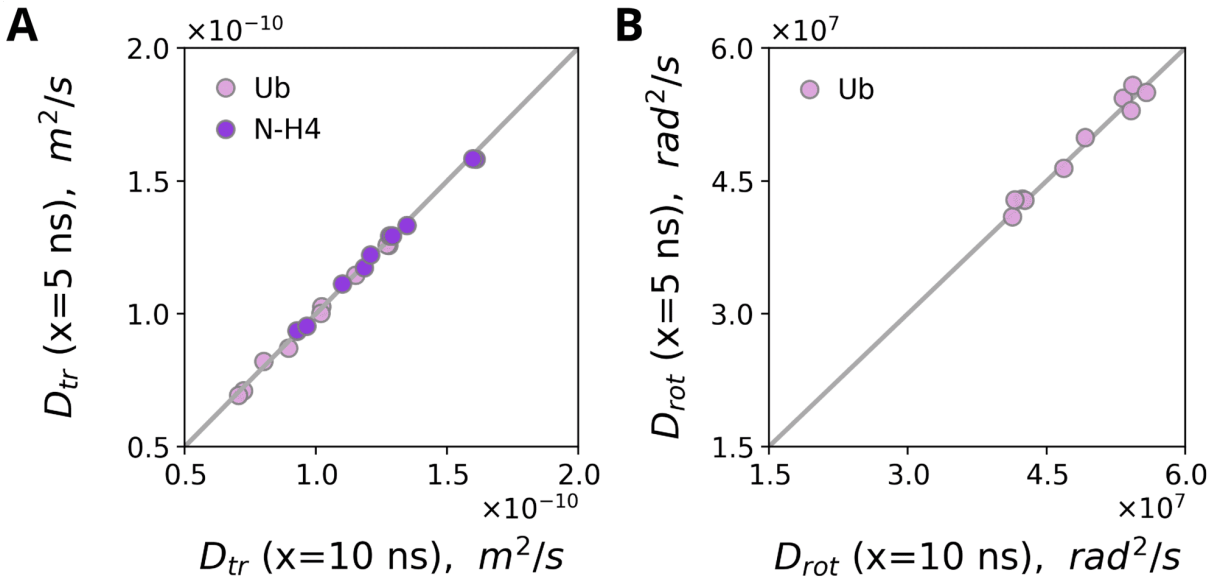


Figure 1.10 — (A) Correlation between the D_{tr} values obtained from the processing of $x = 5$ ns and $x = 10$ ns sets of trajectories. The results are from the fragmented medium- and large-box trajectories of N-H4 and Ub (purple and pink symbols, respectively), including both NPT and NVE simulations. (B) Correlation between the D_{rot} values obtained from the processing of $x = 5$ ns and $x = 10$ ns sets of trajectories for Ub. The results are from the fragmented medium- and large-box trajectories of Ub, including both NPT and NVE simulations.

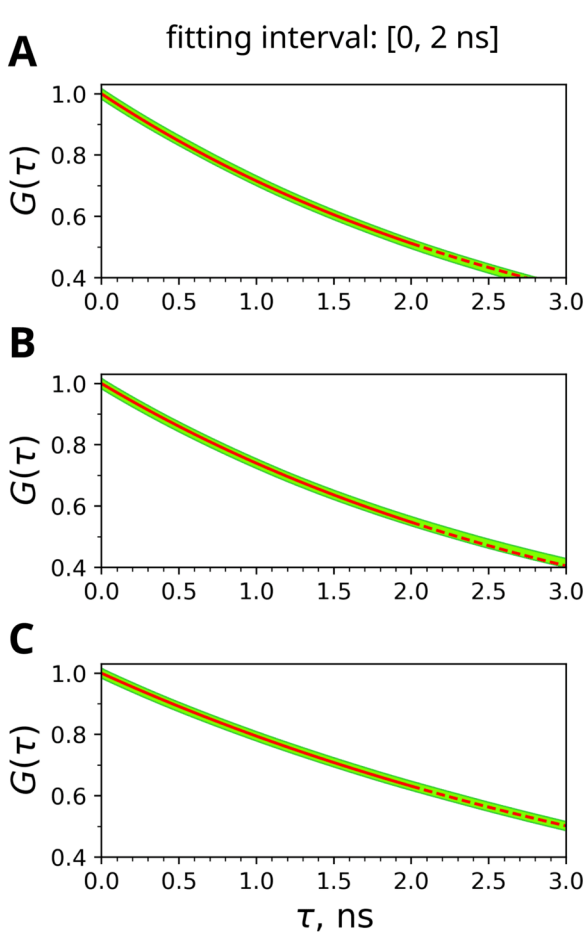


Figure 1.11 — Extracting D_{rot} values from MD simulations of ubiquitin. The $G(\tau)$ time-correlation curves obtained as described in section 1.3 (broad green lines) have been fitted using a monoexponential fitting function $\exp \cdot(-6 \cdot D_{rot}\tau)$. The best fit is shown as a solid red line over the τ interval where the fitting has been performed, [0, 2 ns]; it is shown as a dashed line outside this interval. (A) fragmented trajectory ($x = 10$ ns) in TIP4P-Ew water, large box, NPT ensemble; (B) continuous trajectory in TIP4P-Ew water, small box, NVE ensemble; (C) continuous trajectory in TIP4P-D water, small box, NPT ensemble.

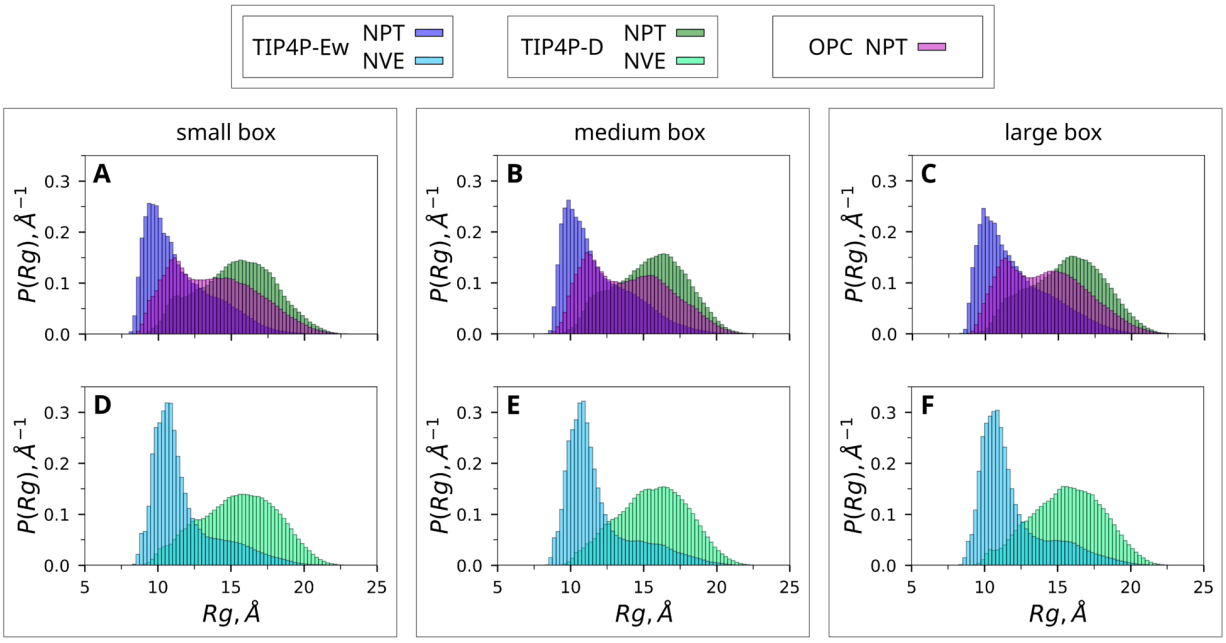


Figure 1.12 — Gyration radius distributions $P(R_g)$ for the N-H4 simulations in TIP4P-Ew, TIP4P-D and OPC water using the NPT ensemble with Bussi thermostat: (A) small-box simulations (reproduced Figure 1.3); (B) medium-box simulations; (C) large-box simulations. Also, gyration radius distributions $P(R_g)$ for the N-H4 simulations in TIP4P-Ew and TIP4P-D water using the NVE ensemble: (D) small-box simulations; (E) medium-box simulations; (F) large-box simulations. The color-coding is described in the figure legend.

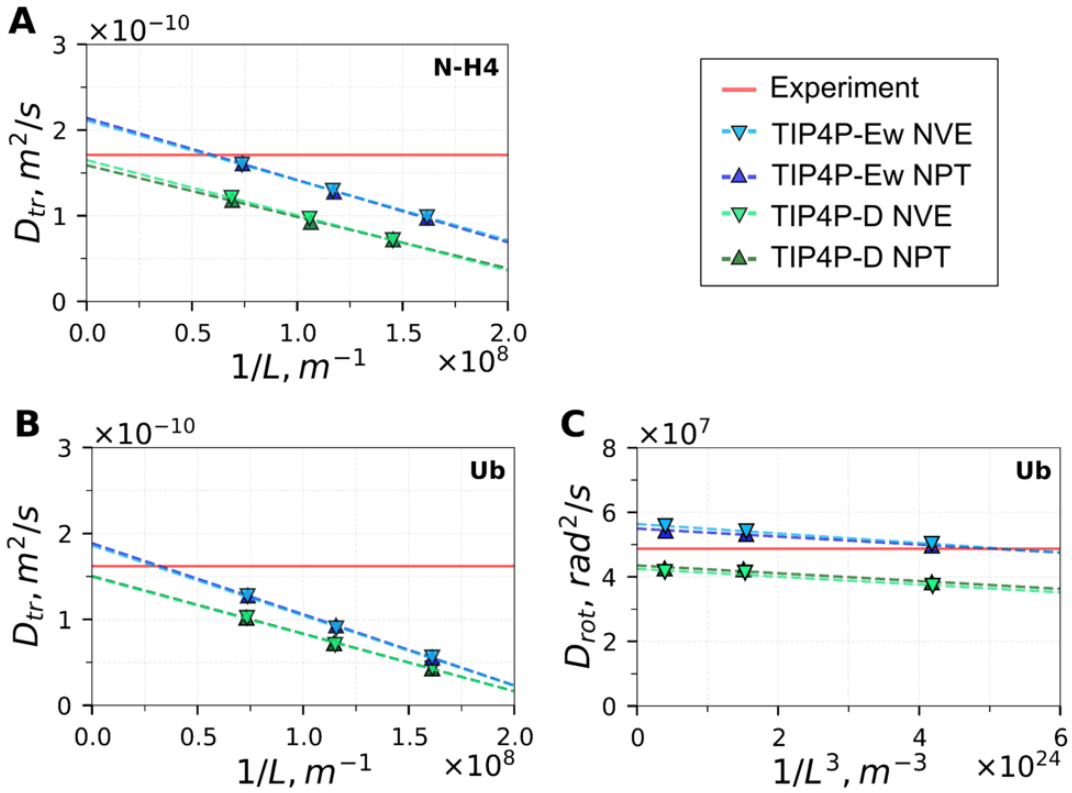


Figure 1.13 — Determination of diffusion coefficients for (A) N-H4 and (B, C) Ub from the series of MD simulations employing TIP4P-Ew and TIP4P-D water models. The results are from the NPT simulations (dark blue and dark green symbols, respectively; reproduced from Figure 1.4), as well as the NVE simulations (light blue and light green symbols). The protocol to process the data is the same as in Figure 1.4.

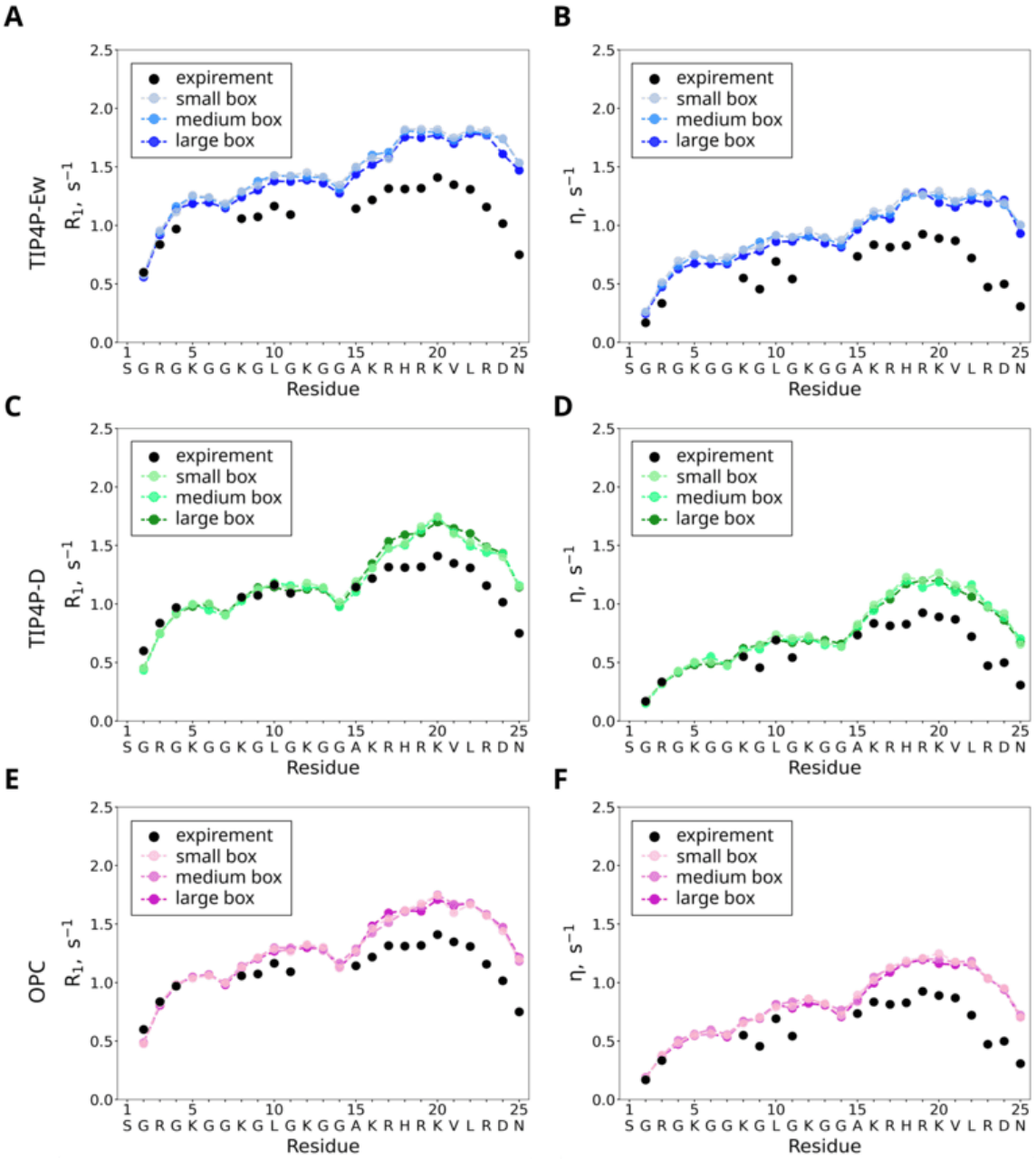


Figure 1.14 — Experimental and simulated ^{15}N relaxation rates R_1 and η in the N-H4 peptide, including the simulations in (A, B) TIP4P-Ew, (C, D) TIP4P-D and (E, F) OPC water. The results from the NPT simulations in the boxes of different size are shown with different color shades as indicated in the figure legends.

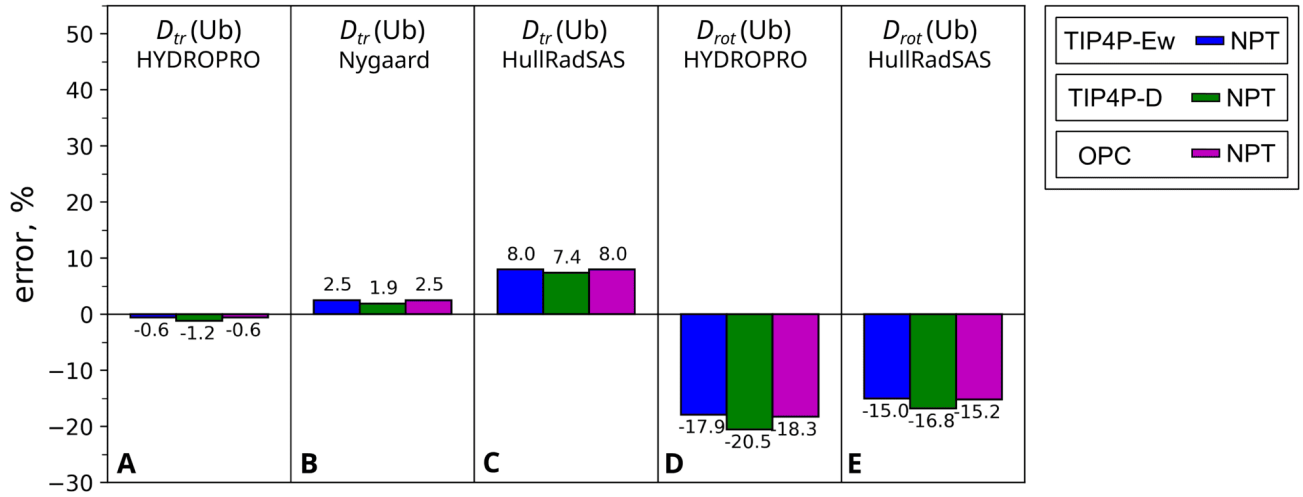


Figure 1.15 — The differences between the calculated and experimental values of diffusion coefficients for ubiquitin. The calculations use 2000 frames from the small-box NPT trajectories of Ub in TIP4P-Ew, TIP4P-D and OPC water (blue, green and magenta bars, respectively). The calculated D_{tr} values are from the following computational tools: (A) HYDROPRO program; (B) empirical formula by Nygaard, which emulates the HYDROPRO results; (C) HullRadSAS program. In the calculations B and C, the calculated values of R_h were converted into D_{tr} by means of the Einstein-Stokes formula (1.3). The alternative conversion method using the data from the small reference molecule, formula (1.4), produces similar results (not shown). The calculated D_{rot} values are from (D) HYDROPRO program and (E) HullRadSAS program.

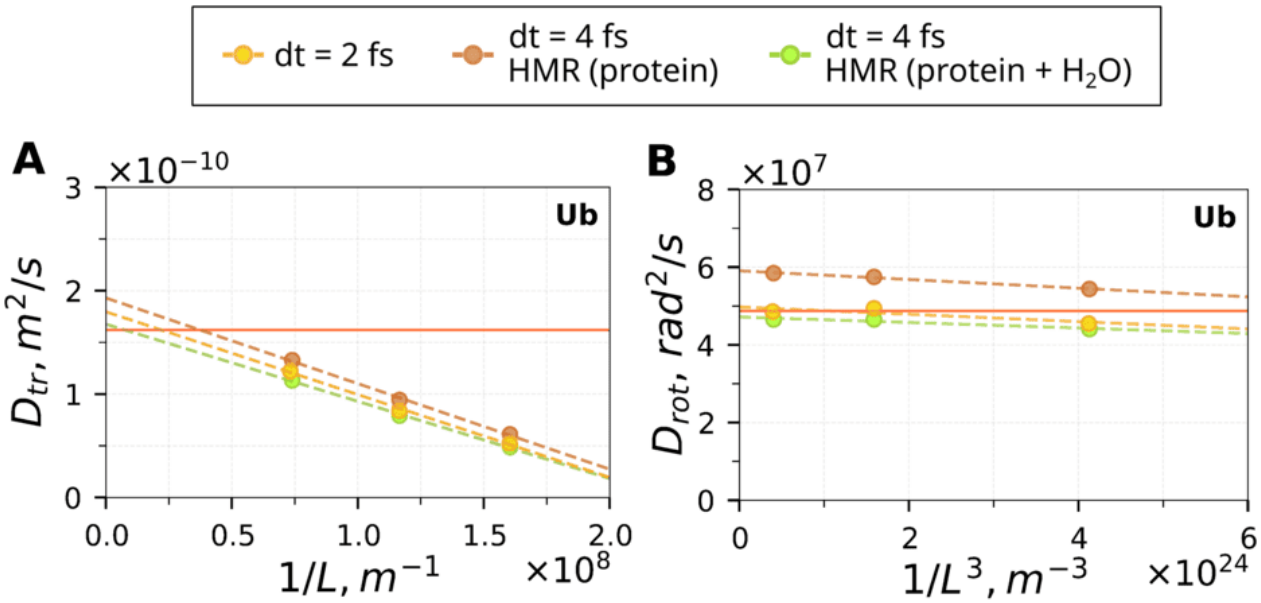


Figure 1.16 — Determination of (A) translational and (B) rotational diffusion coefficients of Ub from MD simulations in the OPC water using extended integration timestep: $\Delta t = 2$ fs (orange symbols), $\Delta t = 4$ fs with the HMR scheme applied to Ub only (brown symbols), and $\Delta t = 4$ fs with the HMR scheme applied to Ub and water (light green symbols).

Table 4 — Summary of MD trajectories of N-H4 recorded in this study

Protein	Water model	Ensemble	Box	Shell thickness	Modeling
N-H4	TIP4P-Ew	NPT	small	12 Å	1 × 5000 ns
			medium	24 Å	500 × 10 ns
			large	48 Å	500 × 10 ns
		NVE	small	12 Å	1 × 5000 ns
			medium	24 Å	500 × 10 ns
			large	48 Å	500 × 10 ns
	TIP4P-D	NPT	small	15 Å	1 × 5000 ns
			medium	24 Å	500 × 10 ns
			large	48 Å	500 × 10 ns
		NVE	small	15 Å	1 × 5000 ns
			medium	24 Å	500 × 10 ns
			large	48 Å	500 × 10 ns
	OPC	NPT	small	15 Å	1 × 5000 ns
			medium	24 Å	500 × 10 ns
			large	48 Å	500 × 10 ns

Table 5 — Summary of MD trajectories of Ub recorded in this study

Protein	Water model	Ensemble	Box	Shell thickness	Modeling
Ub	TIP4P-Ew	NPT	small	12 Å	1 × 2000 ns
			medium	24 Å	200 × 10 ns
			large	48 Å	200 × 10 ns
		NVE	small	12 Å	1 × 2000 ns
			medium	24 Å	200 × 10 ns
			large	48 Å	200 × 10 ns
	TIP4P-D	NPT	small	12 Å	1 × 2000 ns
			medium	24 Å	200 × 10 ns
			large	48 Å	200 × 10 ns
		NVE	small	12 Å	1 × 2000 ns
			medium	24 Å	200 × 10 ns
			large	48 Å	200 × 10 ns
	OPC	NPT	small	12 Å	1 × 2000 ns
			medium	24 Å	200 × 10 ns
			large	48 Å	200 × 10 ns

Table 6 — Experimental and MD-predicted diffusion coefficients for N-H4 and Ub. The MD predictions are from the direct MSD-based interpretation scheme, where per-box D_{tr} and D_{rot} values are extrapolated to the limit of an infinitely large simulation box (illustrated in Figure 1.4 for the NPT simulations and in Figure 1.13 for the NVE and NPT simulations). The errors of the experimentally determined D_{tr} values are from covariance matrix analyses [158], as performed by curve_fit routine in ScyPy [144] and reported by the DDfit server. The error of the experimentally determined D_{rot} value is from the original report [103]. The uncertainties in the MD-predicted D_{tr} and D_{rot} values are from the jackknife procedure whereby we randomly discard 20 % of the MD data (see section 1.3)

Protein	Diffusion coefficient	Experimental values	Simulations				
			TIP4P-Ew		TIP4P-D		OPC
			NPT	NVE	NPT	NVE	NPT
N-H4	D_{tr} (10^{-10} m^2/s)	1.71 ± 0.01	2.14 ± 0.02	2.11 ± 0.02	1.59 ± 0.01	1.65 ± 0.03	1.81 ± 0.02
Ub	D_{tr} (10^{-10} m^2/s)	1.62 ± 0.04	1.89 ± 0.02	1.86 ± 0.02	1.50 ± 0.02	1.50 ± 0.02	1.69 ± 0.02
	D_{rot} (10^{7} rad^2/s)	4.87 ± 0.02	5.50 ± 0.04	5.64 ± 0.05	4.35 ± 0.04	4.25 ± 0.04	4.94 ± 0.05

Chapter 2. Paramagnetic relaxation enhancements: investigation of the histone H4 tail in the nucleosome core particle by MD modelling and NMR spectroscopy

2.1 Introduction

Chromatin is a dynamic protein-DNA complex, which acts to compact eukaryotic genomes and regulate DNA accessibility. The basic repeat unit of chromatin, the nucleosome core particle (NCP), contains ∼147 base pairs (bps) of DNA double helix wrapped ∼1.7 times around a histone octamer protein complex composed of two copies each of histones H2A, H2B, H3 and H4. Each histone contains a structured globular core domain and a positively-charged ∼15-35 amino acid (aa) residue N-terminal tail domain. Early experiments using trypsin digestion and nuclear magnetic resonance (NMR) spectroscopy provided initial evidence that histone N-terminal tails are exposed on the outside of nucleosomes and highly flexible [159; 160], and these insights were confirmed by the first high-resolution X-ray crystal structure of the NCP [161]. Indeed, in this and subsequent X-ray (as well as cryoelectron microscopy) NCP structures electron density is typically missing for extreme N-terminal residues, while for residues further downstream the electron density is usually weak and disjointed. In some cases, certain histone tail fragments can be resolved due to their interactions with neighboring nucleosomes within the crystal lattice allowing for the tails to be partially reconstructed (e.g., Figure 2.1 (C)) [161; 162].

The fact that histone tail domains are conformationally disordered, dynamic and project outward into the solvent is key to their functional role, where they provide «mooring lines» for hundreds of chromatin-associated proteins (CAPs). The latter regulate chromatin assembly and remodeling, and serve as accessory proteins to activate or repress transcription [163]. Notably, histone tails are also processed by histone-modifying enzymes (HMEs), resulting in addition or removal of various post-translational modifications (PTMs). Among the most common PTMs are lysine and arginine methylation, as well as (charge-altering) lysine acetylation [164]. The pattern of these and other modifications, referred to as the «histone code» [165], serves to tune the affinity

of histone tails for different CAPs, as well as their interactions with proximal nucleosomes in chromatin [166]. In summary, histone tails are a focal point of an extremely vast and complex dynamic network, of which we have only a relatively limited understanding at the atomistic level.

In the present study we pursue detailed characterization of the structural ensemble of the dynamically disordered histone H4 N-terminal tails in nucleosomes. Previous studies employing solution and solid-state NMR methods have found that ∼15-20 N-terminal H4 residues are largely dynamically disordered in soluble nucleosomes [97], and that this dynamic disorder persists in compacted nucleosomes [167] and in highly-condensed large nucleosome arrays [98]. Due to its high positive charge, the H4 N-terminal tail is expected to generally localize near and interact with negatively charged nucleosomal DNA (nDNA). However, quantitative measurements of residue-specific NMR spin relaxation rates for H4 tails in nucleosomes coupled with microsecond time-scale molecular dynamics (MD) simulations revealed that electrostatic and hydrogen bonding interactions between H4 tail residues and nDNA are highly transient and dynamic even at low ionic strength [50], following the «fuzzy complex» paradigm [50; 168]. While this previous study of H4 tails in nucleosomes yielded valuable insights about the nature of nDNAH4 tail interactions, the experimental NMR methods used were not suited to directly probe proximities of H4 tail residues to different regions of the nucleosome or the extent to which the H4 N-terminal tail is sequestered by nDNA. These aspects, therefore, remain largely unexplored, although they are potentially important for understanding H4 tail accessibility to CAPs (and HMEs) and its availability to form internucleosomal connections [166], as well as the recently reported cross-talk between H4 and H3 tails [101; 169].

As noted above, high-resolution structural techniques such as X-ray crystallography or cryo-EM are generally unable to localize the disordered histone tail domains in nucleosomes. Instead, this problem can be uniquely addressed by multidimensional NMR spectroscopy [50; 97; 98; 100; 101; 167; 168; 170—180] — in particular, by measurements of residue-specific proton paramagnetic relaxation enhancements (PREs), which report on distances between individual backbone amides and covalent paramagnetic tags attached at specific protein sites and are significant over distances up to ∼20-30 Å. For structured biomolecules in solution, the PRE-based restraints can be

used to build or refine structural models [181—183] and, indeed, studies along these lines have been reported for nucleosome particles and their complexes with CAPs [184—186]. This approach relies on the well-known Solomon-Bloembergen equation [187] that represents PRE rates as a product of distance- and dynamics-dependent terms, where the latter can be determined from heteronuclear relaxation measurements [188] and the former can be converted into structural restraints. Additionally, methods have been developed to account for the conformational flexibility of paramagnetic tags [183; 189]. In contrast, for dynamically disordered protein domains, such as histone N-terminal tails in nucleosomes, the distance- and dynamics-related variables cannot be readily deconvoluted. In these cases, the measured PRE rates are often interpreted in a qualitative manner using one of the empirical random-coil models as a point of reference [190—192]. Otherwise, it is common to make certain simplifying assumptions about the dynamics of the disordered peptide chain [193—195]. Using these assumptions, one can employ PRE data to construct structural models of disordered proteins, typically in a form of multi-member conformational ensembles [196—198]. In addition, specialized methods have been developed to facilitate the extraction of distance information and circumvent the dependence of PREs on dynamics [199; 200].

In our earlier studies of disordered proteins, we have pointed out that PREs in such systems can be rigorously calculated from MD trajectories [13; 201]. Historically, the use of MD simulations to analyze PRE data has been hampered by two factors: the short length of MD trajectories and poor performance of MD force fields in modeling of disordered polypeptides. Both of these factors are no longer as limiting. With the advent of GPU computing it has become possible to routinely record microsecond trajectories of fully or partially disordered systems in large simulation cells [202]. At the same time, a new generation of force fields has been developed (notably including new water models) with the intent to model both structured and disordered proteins [203; 204]. Here we take advantage of these recent advances to show that long MD simulations of the nucleosome core particle can broadly reproduce the experimental PREs measured for histone H4 N-terminal tails in nucleosome samples carrying paramagnetic nitroxide tags attached to different histone H3 sites. Altogether, our results provide a realistic model for the conformational ensemble and interactions of the dynamically disordered H4 tails in nucleosomes.

2.2 Results and discussion

2.2.1 Spin labeled nucleosome samples

Four different paramagnetically labeled NCP samples reconstituted with recombinant *Xenopus laevis* histones and the 147-bp Widom 601 DNA motif [205] were used in this study. Specifically, as described in detail in section 2.3, we employed a histone H3 construct containing the well-studied C110A mutation [97; 98] that removes the lone native cysteine residue while preserving the native histone structure and, starting with this construct, generated the following single-cysteine mutants of H3: K36C, L65C, K79C, and Q125C (see Figure 2.1 (A)). Each H3 mutant was combined with ^{15}N-labeled histone H4 along with unlabeled histones H2A and H2B to produce the histone octamer complex, which was then incubated with excess MTSL nitroxide spin label reagent resulting in the tagging of histone H3 with the non-native paramagnetic side-chain R1 (see Figure 2.1 (B)) at the unique cysteine site [206; 207] and subsequently reconstituted with the Widom 601 DNA to generate the NCP sample used for the NMR studies. For brevity, the spin labeled NCP samples are referred to as 36R1, 65R1, 79R1 and 125R1.

Note that the specific H3 mutation sites selected for this study all correspond to solvent-exposed residues located near nDNA in the coil region preceding helix αN (K36), the ends of helix α1 (L65 and K79) and the center of helix α3 (Q125) to avoid any significant perturbations of the NCP structure and assembly. Potential structural perturbations were further minimized by the use of a tagging protocol which involves spin labeling of the properly assembled histone octamer complex as opposed to individual H3 molecules, with sample fidelity confirmed by mass spectrometry and gel electrophoresis (see Figure 2.3 and Figure 2.4). Finally, the locations of paramagnetic tag sites were selected to be within or near the structured globular domain of H3 in order to explore the H4 tail conformational dynamics relative to multiple well-defined reference points within the NCP structure (see Figure 2.1 (C)). These sites were also chosen to be ∼20-60 Å from one another to provide nonredundant data reporting on the spatial distribution of the H4 tail.

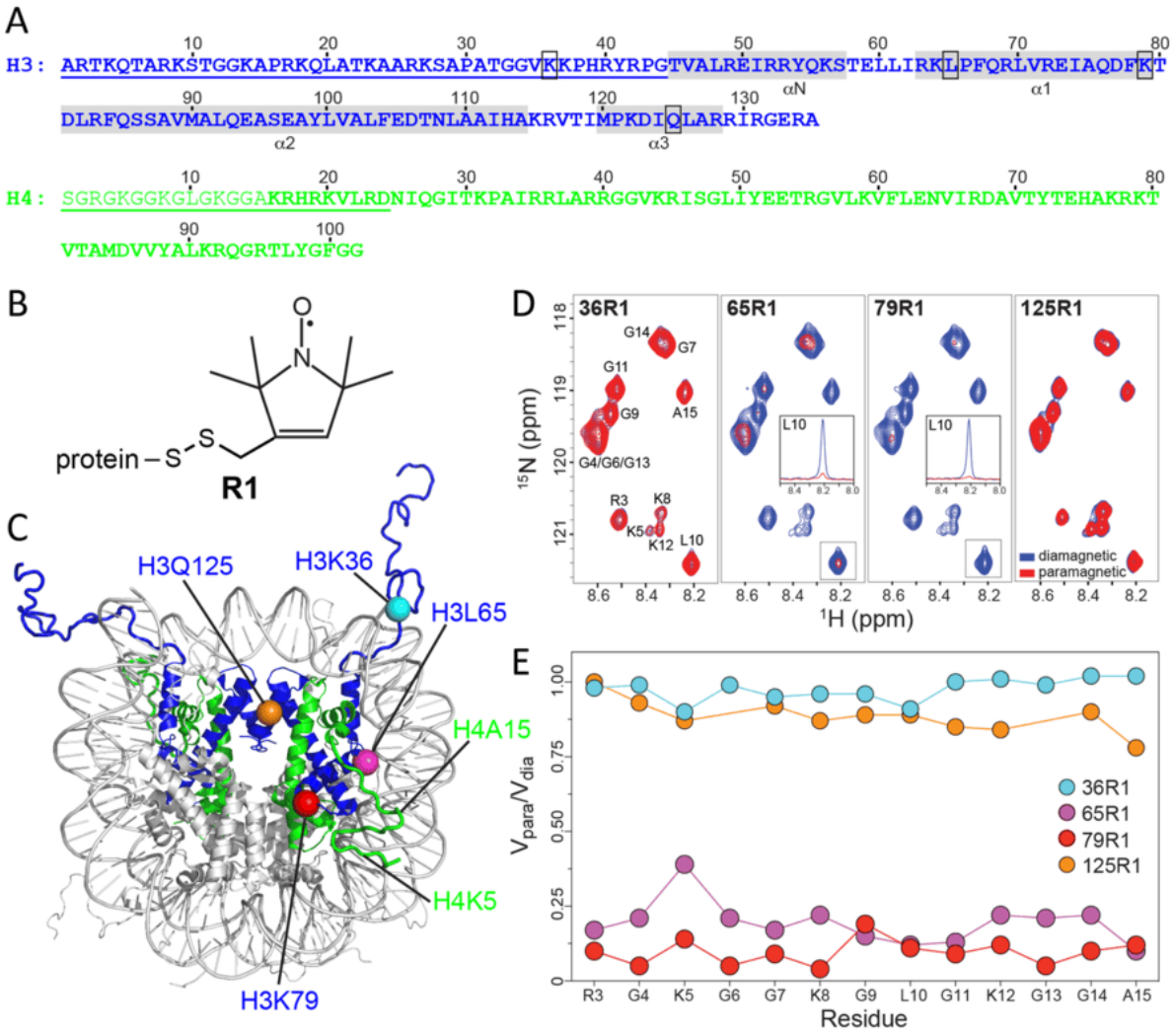


Figure 2.1 — Measurement of paramagnetic relaxation enhancements for H4 histone tail residues in nucleosomes. (A) Amino acid sequences of *Xenopus laevis* histones H3(C110A) and H4. Underlined residues are relatively unstructured and associated with low electron densities and high B factors in the NCP X-ray crystal structure (PDB entry 1KX5) [162]. The H3 core helices are indicated by gray rectangles and labeled underneath; the spin label incorporation sites are denoted by black boxes around the relevant residues. The most flexible H4 tail residues, up to and including A15, are typeset in thin font. (B) Schematic representation of a paramagnetic nitroxide spin label (MTSL) conjugated to a cysteine residue in a protein. The resulting side-chain is termed R1 [206; 207]. (C) Crystal structure of the nucleosome core particle (PDB entry 1KX5) with histone H3 colored blue and histone H4 colored green. The spin-labeling sites in histone H3 (C^{α} atoms of the relevant residues) are indicated by colored spheres and labeled for one of the H3 copies. (D) ^{15}N-$^{1}H^{N}$ HSQC NMR spectra of H4 N-terminal tail in nucleosomes reconstituted with spin-labeled histone H3 and ^{15}N-labeled H4. Shown in each panel is a superposition of paramagnetic (red) and

reference diamagnetic (blue) spectra, plotted at the same contour level, for each NCP sample (as identified at the top of the panel); the reference diamagnetic samples were generated by addition of sodium ascorbate to the corresponding paramagnetic nitroxide spin labeled sample. Resonances for A15 and all glycine residues are aliased in the ^{15}N dimension. Representative one-dimensional ^{1}H traces are shown in the insets for residue L10 in 65R1 and 79R1 samples. (E) Ratios of peak volumes for H4 N-terminal tail residues in HSQC NMR spectra recorded for paramagnetic and diamagnetic NCP samples, V_{para}/V_{dia}, used to determine residue-specific PREs (see Table 7). Given the high sensitivity of the HSQC NMR spectra, the uncertainties in V_{para}/V_{dia} ratios were 0.01 or smaller for virtually all residues in the different samples studied and within the size of the symbols used to depict the data. For clarity the error bars were therefore omitted from the plot. The color coding that identifies the four spin-labeled NCP samples is shown in the figure legend and corresponds to that used in panel (C).

2.2.2 Measurements of PREs in spin labeled nucleosomes

The PRE effects for histone H4 tail residues induced by nitroxide tags attached to histone H3 were quantified by recording for each spin labeled NCP sample a pair of ^{15}N-^{1}H^{N} heteronuclear single quantum coherence (HSQC) NMR spectra: one for the paramagnetic nitroxide spin labeled sample and the other for the same sample but with the nitroxide moiety reduced to diamagnetic hydroxylamine by addition of excess sodium ascorbate (see section 2.3).

The HSQC NMR spectra of nucleosomes reconstituted with ^{15}N-H4 (Figure 2.1 (D)) feature a limited number of resonances corresponding to the most flexible H4 tail residues (aa ∼1-15; note that NMR signals for the two N-terminal residues are not detected due to amide proton exchange [97]), with only one set of peaks observed in the spectra, which indicates that there are no major structural and dynamic differences between the two H4 N-terminal tails within the NCP. The remaining histone H4 residues reorient relatively slowly – including those in the histone core with dynamics dominated by the

slow overall tumbling of the nucleosome particle (∼160 ns [50]) – which leads to severe spectral line broadening and prevents their observation.

Superposition of the paramagnetic and corresponding reference diamagnetic NCP spectra (Figure 2.1 (D)) shows that presence of paramagnetic tags in H3 positions 65 and 79 causes dramatic attenuation of spectral resonances for most of the H4 tail residues. On the other hand, spin labels in H3 positions 36 and 125 cause only relatively minor reduction in H4 tail peak intensities. Qualitatively, this can be readily understood given that the former H3 sites are located in relatively close proximity to the attachment point of the H4 N-terminal tail, whereas the latter two are located some distance away, including residue 36 which is separated from the H4 tail origin by the nDNA helix. For each NCP sample, the ratio of peak volumes in the paramagnetic and diamagnetic HSQC NMR spectra, V_{para}/V_{dia}, is plotted in Figure 2.1 (E) as a function of H4 residue number, and was used to evaluate the residue-specific PREs (see section 2.3 and Table 7) as described in detail previously [201].

2.2.3 Calculation of PREs from MD trajectories

The experimental PRE data were used to validate two different MD models of the nucleosome core particle. The first one is our own 2-μs NCP trajectory recorded under Amber ff14SB force field [67] using the TIP4P-D water model [18]. Previously, we have shown that this trajectory successfully reproduces ^{15}N relaxation data for the H4 N-terminal tail in nucleosomes [50]. The second one is a collection of NCP trajectories with the net length of 41 μs [102], which has been recorded in the Panchenko laboratory using the same Amber ff14SB force field together with the OPC water model [208]. While TIP4P-D was originally designed to improve the modeling of disordered polypeptides (such as H4 tails), OPC was initially devised as a general-purpose model and later shown to perform well for disordered systems [45]. The methodology used to calculate PREs from MD data is described in detail in the section 2.3. Importantly, the algorithm employs rigorous Redfield-theory expressions [201], which fully account for the relative motion of ^{1}H^{N} spins (residing on the H4 tail) and the paramagnetic center (residing on the core histone H3). This motion mainly stems from the

conformational dynamics of the histone H4 tail and can be thought of as restricted diffusion; it efficiently modulates both the orientation of the interspin vector and its length (i.e., the distance between the two spins). The distinctive correlation functions resulting from these dynamics are illustrated in Figure 2.5.

Before we compare the MD-based predictions to the experimental results, we note that for a given spin-labeled sample there is relatively little residue-to-residue variation in the measured PREs over the entire length of the N-terminal tail (c.f., Figure 2.1 (E) and Table 7). Considering limited convergence of the MD-based residue-specific PRE calculations (discussed in more detail below), it is unlikely that these small residue-to-residue PRE variations can be reproduced. Therefore, we have elected instead to focus on the tail-averaged experimental PREs for each of the four spin-labeled NCP samples and compare these with the similarly averaged calculated PREs, which are less susceptible to the convergence issues.

2.2.4 Comparison of experimental and calculated PREs

Figure 2.2 (A, B) show the comparison of experimental and MD-derived PREs for simulations employing TIP4P-D and OPC water, respectively. In these plots, coral and green bars indicate the experimentally measured and calculated values, respectively. Note that the calculated PREs shown in these graphs constitute strictly a prediction, with no tunable parameters of any kind. Clearly, both sets of calculated PREs, based on TIP4P-D and OPC trajectories, capture the qualitative behavior as observed in the experiment, with much higher paramagnetic rates predicted for H3 79R1 and 65R1 samples than for 125R1 and 36R1 samples. While the MD simulations fail to predict the PREs with quantitative accuracy it is encouraging that TIP4P-D results and OPC results tend to «bracket» the experimentally determined PRE values. For example, the TIP4P-D simulation correctly recovers experimental 36R1 result, while the OPC simulation produces an overestimated value. Further, the TIP4P-D simulation somewhat underestimates the 65R1 result, while the OPC simulation significantly overestimates it. Finally, the TIP4P-D simulation substantially underestimates the average PRE rate in the 79R1 sample, whereas the OPC

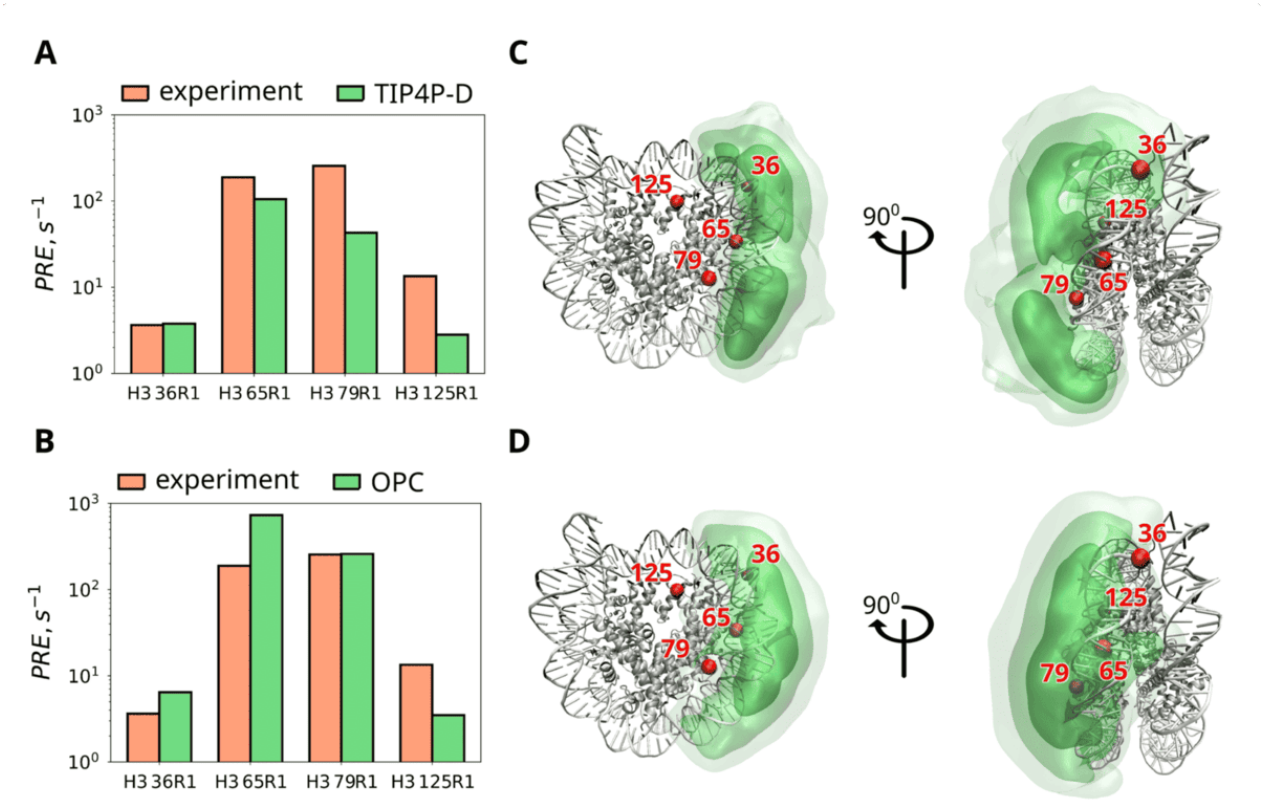


Figure 2.2 — Comparison of experimental and MD-derived paramagnetic relaxation enhancements. (A, B) Experimental PREs averaged over histone H4 residues 3-15 and over two copies of H4 in nucleosomes (coral bars) and the corresponding MD predictions (green bars). The MD results are from TIP4P-D and OPC simulations (panels A and B, respectively). (C, D) The localization of histone H4 tail residues 1-15 in TIP4P-D and OPC simulations (panels C and D, respectively). The images are based on the coordinate set 3LZ0 [209]. In constructing these images, we took advantage of the 2-fold pseudo-symmetry of the NCP such as to combine the data from the two H4 tails. Shown is the density of C^{α} atoms for H4 residues 1-15; the isosurfaces containing 50 %, 90 % and 99 % of the integral density are colored dark, medium and light green, respectively. The nitroxide tag attachment sites in histone H3 are indicated by spheres and labeled (c.f., Figure 2.1 (C)).

simulation provides a good match. Based on these observations, we conclude that the two MD simulations taken together offer a broadly correct model for the PRE effects in spin-labeled nucleosomes.

It is worthwhile to consider the possible reasons for the observed lack of quantitative agreement between experimental and predicted PREs in Figure 2.2.

Almost certainly this has to do with the lack of convergence for the MD-based PRE calculations in the system, which involves 25-residue-long disordered H4 tail probing a large swath of the nucleosome surface. Indeed, the calculated PREs show an exceedingly steep dependence on the interspin distance (see equation (2.3)). As a consequence, even a brief close encounter of the H4 tail with the paramagnetic label can have a dramatic impact on the calculated PREs. In other words, MD-based calculations of residue-specific PREs are extremely demanding in terms of convergence [198]. This idea is further illustrated in Figure 2.6, where the comparison is drawn to four different subsets of the OPC trajectories.

Acknowledging this limitation, we nevertheless note that the results in Figure 2.2A and 2B provide a broad semi-quantitative validation for the MD models at hand. Of special interest to us, these models encode the information about the localization of the histone H4 tails in nucleosomes. This is visualized in Figures 2.2 C and D in a form of spatial density distribution for H4 tail residues 1-15 (volumes shaded with different shades of green). Inspection of these plots confirms that these H4 residues are significantly delocalized, as expected for the dynamically disordered tail. At the same time, the H4 tail clearly gravitates toward the nucleosomal DNA, in agreement with the notion previously discussed by us and others that histone tails engage in fuzzy interactions with nDNA [50; 168].

The results in Figure 2.2 C and D show that H3 labels 65R1 and 79R1 are indeed centrally located with respect to the H4 tail «cloud». On the other hand, 36R1 and 125R1 are located peripherally at a distance from the H4 tail as seen in the side and front views of the NCP, respectively. Of interest, different subsets of the OPC trajectories show some variability in terms of the H4 tail density distribution (see Figure 2.6). Some of these models show the density distribution that is somewhat more extended toward the 125R1 label and, accordingly, predict higher PRE rates in this sample (more in line with the experimental PRE data), suggesting that the details of the density distribution can be further elucidated based on longer trajectories and additional experimental data.

While the MD-based interpretation of PRE data is currently faced with certain technical challenges, including sub-par convergence in calculation of residue-specific PREs, it constitutes a general and powerful strategy the accuracy of which is only expected to improve in the coming years with continuing rapid

advances in simulation length and quality of water models. In contrast, the more traditional static ensemble-based approaches are faced with several inherent limitations in dealing with PRE data as discussed below [198].

First, it is noteworthy that such ensemble-based approaches have no knowledge of the residue-specific correlation times, τ_c, which are necessary to calculate PRE rates but are challenging to determine experimentally. For histone H4 tails in nucleosome, MD simulations indicate that these τ_c values vary from one residue to another and can be significantly shorter than the tumbling time of the entire nucleosome particle, τ_{rot}; furthermore, τ_c cannot be reliably reconstructed from ^{15}N relaxation data (see Figure 2.5). Therefore, when constructing a conformational ensemble one either needs to make assumptions about τ_c or otherwise use it as a tuning parameter.

Second, an earlier study by Ganguly and Chen [198] concluded that PRE-restrained ensemble models of disordered proteins tend to be ill-defined. Specifically, due to the r^{-6} weighting associated with PREs, only one or a few ensemble members are typically needed to satisfy the experimental PREs and the rest of the ensemble could be essentially unrestrained. As such the resulting ensemble may not necessarily be a realistic representation of the underlying disordered state.

These observations also apply to the system in our study as illustrated in Figure 2.7 and Figure 2.7. In Figure 2.7, we present a minimal ensemble consisting of just two H4 tail conformers which clearly does not represent the conformational diversity of the H4 tail domain in nucleosome core particle yet successfully reproduces the experimental PRE data. Additionally, in Figure 2.8 we show a larger ensemble of 42 H4 tail conformers, where the PRE rates are largely determined by two lightly populated conformers, while the remaining 40 conformers are effectively unrestrained and largely arbitrary.

Finally, we note that validation of the two MD models of the NCP discussed in this paper does not need to be limited to the PRE data. For instance, both TIP4P-D and OPC simulations were successful in reproducing secondary chemical shifts for H4 tail residues 1-15, correctly predicting the random-coil character of this glycine-rich fragment. The TIP4P-D simulation was also successful in predicting ^{15}N relaxation rates for H4 tail residues [50]. In contrast, the OPC simulations seem to overemphasize the attraction between the histone tails and nDNA and thus produce biased predictions for ^{15}N relaxation

rates. Several factors may contribute to this situation, which will be the subject of a separate study. At the same time, the TIP4P-D simulation shows some early signs of nDNA unwrapping which affects the two ends of the DNA superhelix. This behavior likely reflects the known stability issues with this water model [18]. In contrast, the OPC simulations appear to be stable in this regard, although some of the OPC trajectories show evidence of dynamic nDNA fluctuations (see Figure 2.9).

2.3 Materials and Methods

2.3.1 Sample preparation

The pJ201 plasmid containing 32-copies of the 147-bp Widom 601 DNA positioning sequence was transformed into *E. coli* DH5α, amplified in LB rich medium, extracted by using a QIAGEN Plasmid Giga kit, digested with EcoRV-HF (New England Biolabs), and purified by PEG precipitation as described previously [210]. Briefly, 4 M NaCl and 40 % PEG 6000 were added to the EcoRV-digested plasmid in a 0.192:0.346:1 v/v/v ratio. The mixture was incubated on ice for 1 h and then centrifuged (27,000 g, 4 °C, 20 min). The supernatant containing the desired 147-bp Widom 601 DNA sequence was collected and combined with cold ethanol in a 1:2.5 v/v ratio. The precipitated DNA was centrifuged (27,000 g, 4 °C, 40 min), the supernatant decanted, the precipitate air-dried for 10 min and then dissolved in 0.5 x TE buffer (5 mM Tris, 0.5 mM EDTA, pH 7.5).

Histone protein expression and purification. *Xenopus laevis* histones H2A, H2B, H3(C110A) containing a cysteine point mutation at one of the following sites: K36C, L65C, K79C or Q125C, and H4 were overexpressed in *E. coli* BL21 (DE3) pLysS, purified as described previously[211] using gel filtration and ion-exchange chromatography in 7 M urea followed by dialysis against a solution of 2 mM β-mercaptoethanol (BME), and lyophilized. Uniformly ^{15}N-labeled H4 was prepared by using a minimal medium with $^{15}NH_4Cl$ (1 g/L) (Cambridge Isotope Laboratories) as the sole nitrogen source. To express

H3 cysteine mutants for subsequent spin labeling, plasmids were generated by site-directed mutagenesis using the wild-type histone H3 plasmid and Herculase II fusion DNA polymerase (Agilent).

Spin labeling of histone octamer. Histone octamer was prepared by dissolving the four unfolded histone proteins (H2A, H2B, H3 mutant, and ^{15}N-labeled H4) at concentrations of 10 mg/ml in unfolding buffer (7 M guanidine hydrochloride, 20 mM Tris, 10 mM DTT, pH 7.5) in a H2A:H2B:H3:H4 molar ratio of 1.2:1.2:1:1, and then transferred by double dialysis into refolding buffer (1 x TE, 2 M NaCl, 5 mM BME, pH 8.0). After dialysis, the solution was concentrated to a volume of 5 mL by Amicon centrifugal filters (30 kDa cutoff, MilliporeSigma). To ensure complete reduction of the cysteine residue in H3, DTT was added to the solution containing the refolded octamer to a final concentration of 10 mM. The solution was allowed to sit on ice for 30 min, and was then purified by gel filtration chromatography in refolding buffer (without BME and DTT) as described previously [211]. Eluted fractions containing pure octamer were immediately pooled and incubated overnight in the dark at 4 °C with a 20-fold molar excess of the nitroxide spin label reagent (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl)-methanetiosulfonate (MTSL; Toronto Research Chemicals, Inc.). To remove excess MTSL tag, the octamer was transferred into 0.5 x TE buffer and concentrated to a final concentration of 15 mg/mL using an Amicon centrifugal filter. Spin labeling of the octamer was routinely confirmed by MALDI-TOF mass spectrometry (see Figure 2.3), and estimated to be at least ∼80-90 % efficient as assessed by peak deconvolution analysis of spectra before and after tagging.

DNA and histone octamer were combined in a molar ratio of DNA:octamer of 1:0.65, in an aqueous 0.5 x TE, 2 M NaCl and 1 mM benzamidine hydrochloride hydrate (BZA) buffer. The NaCl was removed by double dialysis against 0.5 x TE, 1 mM BZA buffer at 4°C. To separate fully formed nucleosome particles from naked DNA, sucrose gradient centrifugation was performed using a sucrose gradient of 5-30 % in 0.5 x TE buffer. The fractions containing nucleosome particles were combined and the sucrose was removed by buffer exchange into 0.5 x TE buffer using an Amicon centrifugal filter. The formation and purity of nucleosomes was confirmed by an electrophoretic mobility shift assay using native polyacrylamide gel electrophoresis (Figure 2.4).

2.3.2 NMR experiment

Solution NMR experiments on nucleosomes reconstituted with ^{15}N-labeled H4 and MTSL-tagged H3 were recorded at 25 °C using an 850 MHz Bruker Avance III HD spectrometer equipped with a TCI CryoProbe. NMR samples consisted of nucleosomes at concentrations of ∼40-45 µM in aqueous buffer solution containing 5 mM Tris, 0.5 mM EDTA, 100 mM NaCl, 0.1 mM $MgCl_2$ and 5 % D_2O at pH 7.0. For each paramagnetic spin labeled nucleosome sample, following NMR measurements on that sample, a reference diamagnetic sample was generated by addition of a 20-fold molar excess of sodium ascorbate followed by incubation for ∼4-7 h at 4 °C prior to NMR measurements. A simple ^{15}N-^{1}H^{N} HSQC pulse sequence with water flip-back pulses included for water suppression was used to determine residue-specific peak volumes for the different histone H4 tail amide sites in the paramagnetic and diamagnetic samples, and the PREs (see Table 7) were extracted from the peak volume ratio according to the following expression: $V_{para}/V_{dia} = \exp(-PRE \cdot 2\tau_{inept})$ as described in detail previously [201]. The duration of the INEPT elements in the HSQC pulse sequence was 4.56 ms, and the recycle delay was set to 2 s ensuring complete signal recovery for both paramagnetic and diamagnetic samples.

2.3.3 MD modelling

The simulations of NCP in TIP4P-D water (a single 2-µs trajectory) and in OPC water (twenty two trajectories with the net length of 41 µs) have been described previously [50; 102]. Both simulations use the same force field, Amber ff14SB [67], and are generally similar apart from the water model. Below we comment on some of the relevant aspects of these simulations.

First, in our TIP4P-D simulation the initial structure is based on the PDB entry 3LZ0, matching the experimentally investigated system. On the other hand, the study by Peng and co-authors [102] pursued different goals; in their OPC simulations, they used the model based on the PDB entry 1KX5 [162], which has the same *Xenopus laevis* histone core, but a different DNA sequence.

Furthermore, they have replaced the original nDNA with the sequence from human KRAS gene and, in addition, extended the nDNA from both ends by adding 20-bp linker segments. In principle, this type of structure editing can have a perturbing effect on the NCP simulation. Incidentally, in some of the trajectories the outer turn of nDNA is dynamic, but we found no evidence of progressive nDNA unwrapping (Figure 2.9).

Second, TIP4P-D simulations were conducted at the temperature 25 °C in the presence of 100 mM NaCl, which matches the experimental conditions used in this study. On the other hand, OPC simulations were conducted at 36 °C in the presence of 150 mM NaCl. Although elevated temperature and salt concentration should promote the mobility of histone tails, we have observed that H4 tail dynamics are actually slowed down in the OPC simulations compared to TIP4P-D simulations.

Third, the OPC trajectories released by Peng et al. [102] have been sampled with a time step of 1 ns. To investigate the effect of the step size, we reprocessed our TIP4P-D trajectory with the sampling step of 1 ns and compared the results with the original calculations using 1 ps grid. As it turns out, the PREs and ^{15}N R_2 rates remain essentially unchanged when we switch to the sparsely-sampled data (results not shown). On the other hand, ^{15}N R_1 rates prove to be sensitive to the sampling step. This is to be expected given that ^{15}N rates are sensitive to faster forms of motion, in contrast to R_2 rates that are mainly determined by slow motional modes. If the goal is to compute rates, the sampling step of 10 ps can be recommended.

Fourth, both TIP4P-D and OPC simulations employed the widely used Langevin thermostat. We note that Bussi-Parrinello velocity rescaling thermostat [63] is, in fact, better suited to calculate the dynamics-dependent relaxation rates.

MD-based calculations. The paramagnetic contribution to spin-spin relaxation, measurable on per-residue basis for backbone ^{1}H^{N} spins, is given by the following expressions:

$$\Gamma_2 = \frac{1}{20} D^2 (J(0) + 4(J(\omega_H)), \tag{2.1}$$

$$J(\omega) = \int_0^\infty C(\tau) \cos(\omega\tau) d\tau, \tag{2.2}$$

$$C(\tau) = \left\langle \frac{P_2(\cos\theta(t,t+\tau)}{r^3(t)r^3(t+\tau)} \right\rangle \tag{2.3}$$

Here D is the interaction constant, $D = 4.9764 \cdot 10^{-22}$ m^3·s$^-$1 for the MTSL label, $J(\omega)$ is the spectral density, ω_H is the proton Larmor frequency, $C(\tau)$ is the dipolar correlation function, $P_2(x)$ is the second-order Legendre polynomial, $\theta(t,t+\tau)$ is the angle between the vectors $\vec{r}(t)$ and $\vec{r}(t+\tau)$ connecting paramagnetic center with the given ^{1}H^N spin at the points in time$\vec{r}(t)$ and $\vec{r}(t+\tau)$, respectively, with the corresponding vector lengths $r(t)$ and $r(t+\tau)$, and the angular brackets denote averaging over time and over the ensemble (consisting in this case of the two copies of histone H4). Considering the two-fold pseudo-symmetry of the NCP, each ^{1}H^N spin residing on the H4 tails is relaxed by two MTSL labels attached to the two copies of the histone H3, termed «proximal» and «remote»

$$PRE = \Gamma_2^{proximal} + \Gamma_2^{remote} \tag{2.4}$$

To calculate the PRE rates from the MD trajectories, we first overlay all MD frames by superimposing them onto the reference coordinates 3LZ0 via the secondary-structure C$^\alpha$ atoms from the histone core. The relevant correlation functions are then calculated as follows

$$c(\tau) = \frac{1}{N_{total}} \sum_{i=1}^{N_{traj}} \sum_{j=1}^{2} \sum_{k=1}^{N_{timepointpairs}(i)} \frac{P_2(\cos\theta(t,t+\tau)))}{r^3(t)r^3(t+\tau)} \tag{2.5}$$

where the first summation is over all available trajectories representing the system of interest, the second summation is over the two H4 tails in the simulated nucleosome particle, and the third summation is over all timepoint pairs $t, t+\tau$ in the given trajectory (considering a fixed value of τ). The result is normalized by N_{total}, which is the total number of terms arising from all three summations in equation (2.5). When calculating $\Gamma_2^{proximal}$, the second summation in this formula implies that the two vectors $\vec{r}(t)$ are involved: (1) between the first H4 tail and its proximal spin label and (2) between the second H4 tail and its proximal spin label. The remote contribution Γ_2^{remote} is evaluated in the analogous way. Note also that the distinction between the proximal and remote spin labels is based

not only on distance, but also on more fundamental symmetry considerations. Specifically, 180° rotation about the dyad axis interchanges the two proximal labels (and also interchanges the two remote labels).

It is worth stressing that equation (2.5) is perfectly suited to combine the data from the two (presumed equivalent) H4 tails in the simulated NCP particle, as well as the data from multiple MD trajectories modeling NCP dynamics. More generally, this formula performs both ensemble- and time-averaging for the problem at hand. The approach using equation (2.5) is much preferable to a hypothetical alternative, where the PRE rates are computed separately for each of the two tails and each of the many trajectories and then averaged to arrive at the final result. The correlation functions $c(\tau)$ calculated via equation (2.5) are subsequently multiplied by $\exp(-t/\ \tau_{rot})$ to reintroduce the tumbling motion of the nucleosome core particle:

$$C(\tau) = c(\tau)\exp(-t/\ \tau_{rot}) \tag{2.6}$$

The value of τ_{rot}, 163 ns, is from HYDROPRO calculations, as reported previously [50]. The function $C(\tau)$ is then approximated using a multiexponential fitting function, as described previously [50]. This multiexponential representation is further used to evaluate the spectral densities, equation (2.2), and ultimately to calculate the PREs, equation (2.4).

The MD trajectories at hand do not contain an explicit MTSL tag (explicit modeling of the four MTSL-labeled samples is possible [212], but would mean a shorter simulation for each individual sample). Instead, we have used the heavy atoms at the extremity of the corresponding side chain as a proxy for the MTSL paramagnetic center: N^{ζ} atom in K36 or K79, $C^{\delta 1}$ in L65 and $N^{\varepsilon 2}$ in Q125 in the histone protein H3. One can argue that the motion of these native side chains to a certain degree approximates the motion of the MTSL tag. Also bear in mind that MTSL movements is only a minor source of dynamics in this system compared to the motions of the disordered H4 tail. Generally, given the level of convergence in our MD-based PRE calculations, this approximation seems to be justified.

Of further note, equation (2.5) depends not only on reorientation of the spin-spin vector $\vec{r}(t)$, but also on time modulation of the distance $r(t)$ due mainly to the dynamics of the mobile H4 tail. In this sense, the observed effect is reminiscent of translational relaxation. The influence of constrained translational

dynamics on the correlation functions $C(\tau)$ is rather pronounced. For $^1H^N$ spins at the extremity of the mobile H4 tail interacting with a proximal paramagnetic label the correlation function decays on a time scale of tens of nanoseconds, much faster than dictated by the tumbling motion of the NCP (Figure 2.5).

To visualize the location of H4 tails in the NCP (Figure 2.2 and Figure 2.6), we have merged the data from the two histone tails. For this purpose, we first prepared the rotated copy of the reference structure 3LZ0. This was accomplished by superimposing the first copy of each histone, H2A, H2B, H3 and H4, onto the second copy of the same histone and vice versa (the histones were superimposed via C^α atoms from the secondary-structure regions). This transformation corresponds to 180° rotation about the dyad axis. Subsequently, all MD frames have been superimposed onto the rotated structure 3LZ0, thus obtaining a rotated copy of the original trajectory. Finally, we extracted the coordinates of C^α atoms for residues 1-15 in H4 tail (second copy) in the rotated trajectory and transferred them to the original trajectory. In this manner we have generated a reduced trajectory comprised of the original structure 3LZ0, plus the MD-derived coordinates of C^α atoms from the two (co-localized) H4 tails. The coordinates of C^α atoms were used to produce a volumetric density map using the VolMap plugin of the program VMD v.1.9.1 [213] with the following options: resolution 1 Å, Gaussian smoothing factor 3, no weights setting. The maps were visualized in VMD using the Isosurface representation [213].

To generate Figure 2.7, we first formed the pool of conformers by extracting NCP coordinates from the TIP4P-D trajectory using the step of 1 ns. Each copy of the H4 tail was treated as an individual conformer, characterized by its position relative to the paramagnetic labeling sites. For each conformer, we calculated the PRE rates based on equations (2.1) и (2.4), but replacing equation (2.2) with the following simple formula [193]

$$J(\omega) = \frac{1}{r^6}\frac{\tau_c}{1+\omega^2\tau_c^2} \tag{2.7}$$

where r is the conformer-specific distance between a certain amide proton in the H4 tail and the paramagnetic center, while τ_c is the effective correlation time.

As discussed in the text, the standard procedure to construct an experimentally restrained conformational ensemble has no prior knowledge of τ_c. For lack of a better option, we set τ_c equal to $\tau_{rot} = 163$ ns. On a physical level,

this means that the motion of the H4 tail over the surface of the nucleosome core particle is assumed to be slow on a time scale of τ_{rot} (which is a rather arbitrary assumption). It is further assumed that, while loosely anchored to the surface of NCP, the H4 tail remains conformationally labile («fuzzy» behavior), which is a necessary condition to observe spectral signals by solution NMR. After that we have searched for a combination of two conformers that would produce the best agreement between the calculated and the experimental PRE rates (in a sense of Figure 2.2 A). For this purpose, we have used the exhaustive search within the pool of four thousand conformers; for each pair of conformers, we also optimized their relative weights.

To generate Figure 2.7, we first formed the pool of conformers representative of the OPC trajectories of NCP [102] (resampled with the step of 10 ns). Similar to the above, we have identified two conformers that closely reproduce the experimental PRE data, but in doing so we assumed that both conformers are populated at the level of 10 %. In addition, we have also selected 40 conformers that contribute only minimally to the PRE rates; each of these additional conformers was assigned the weight of 2 %.

While Figures 2.7 and 2.8 are based on a collection of MD snapshots, it is also possible to generate relevant conformational ensembles by other means. It should be noted, however, that standard tools, such as Flexible-meccano [117] or IDPConformerGenerator [214], are insufficient for this purpose. Indeed, after a random conformation of the H4 tail is generated, it needs to be «glued» onto the body of the nucleosome core particle; then the resulting model should be evaluated from the standpoint of steric clashes and either accepted, subjected to a remediation procedure, or discarded. Irrespective of the algorithmic details, PRE-restrained ensembles of disordered systems tend to be problematic, as first pointed out by Ganguly and Chen [198] and illustrated by Figures 2.7 and 2.8. The situation can be improved by increasing a number of paramagnetically labeled samples and also by employing additional experimental restraints (e.g. chemical shifts, relaxation rates, etc.).

2.4 Concluding Remarks

Characterization of conformational ensembles for the dynamic histone tail domains in the context of nucleosomes and nucleosome arrays is important for understanding of their interactions and ultimately function. Of particular importance are changes in tail conformation and interactions that arise in response to different PTMs and involvement of chromatin-associated proteins. These problems can be uniquely addressed by probing a range of NMR observables in both solution and solid state as appropriate. In this study, we have demonstrated that measurements of residue-specific PREs in nucleosome samples employing multiple paramagnetic tags should be one of the central elements of such multi-experiment NMR datasets. We have also shown that interpretation of these PRE data based on long MD simulations of nucleosome particles yields valuable insights about the conformational and interaction landscapes of histone tails in these systems. In the future we envision that this strategy can be extended to nucleosome complexes with CAPs and even larger nucleosome assemblies, as well as to other large biomolecular complexes containing functionally important flexible segments. Finally, it is noteworthy that PREs and other types of experimental NMR data can be used to validate MD models and, by extension, to inform efforts in the area of force field development. Collectively, these efforts are expected to contribute to an improved understanding of the functional mechanisms in chromatin.

2.5 Supporting Information

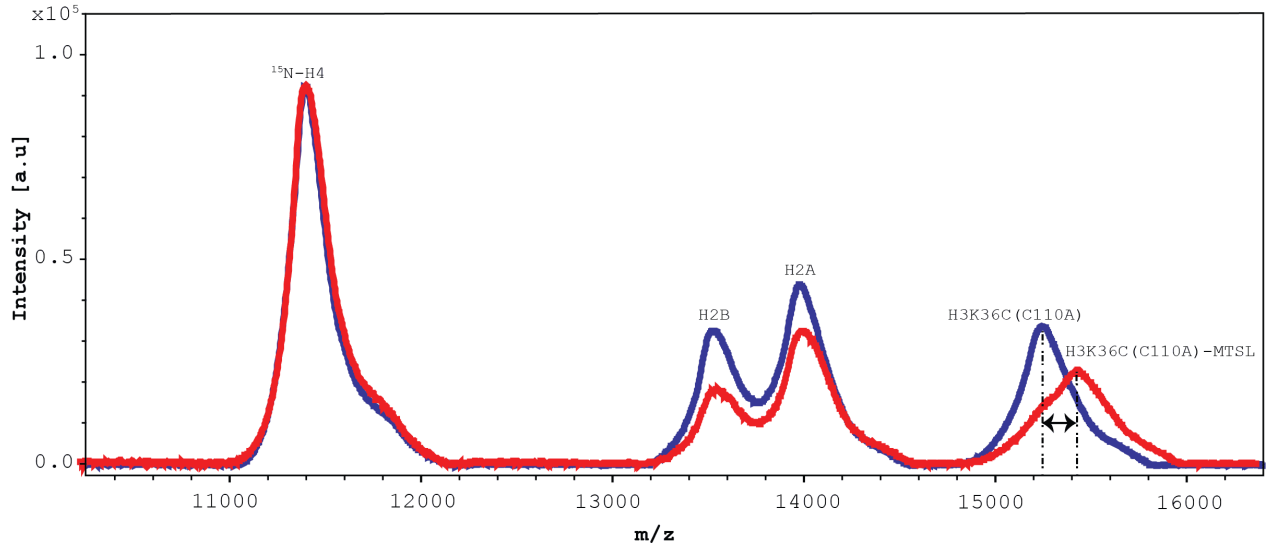


Figure 2.3 — Representative MALDI-TOF mass spectra of histone octamer containing ^{15}N-labeled H4 and H3(C110A) K36C before (blue) and after (red) tagging with MTSL. Observed m/z values are consistent with the expected ∼184 Da increase in molecular mass of H3(C110A) K36C due to tag attachment. The vertical scales in the spectra are set to match the intensities of the peak corresponding to ^{15}N-labeled H4.

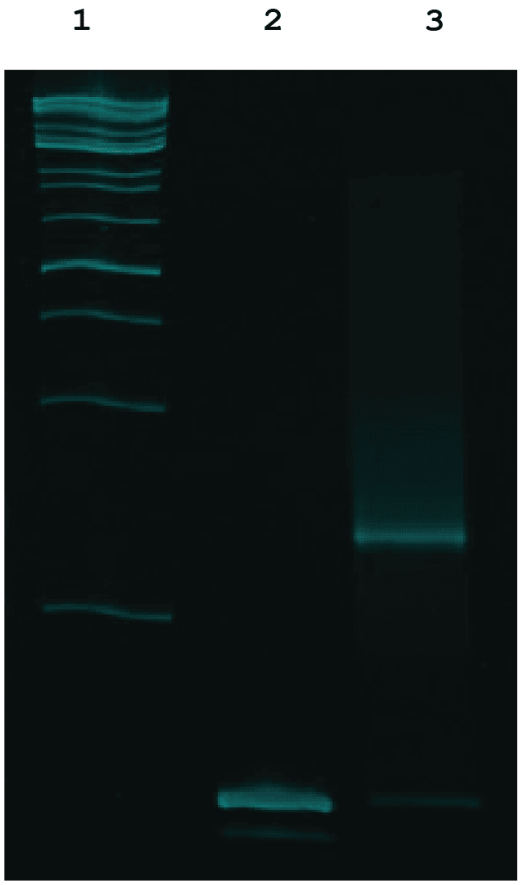


Figure 2.4 — Native polyacrylamide gel (5 % polyacrylamide, 0.3 x TBE running buffer: 27 mM Tris-borate, 0.6 mM EDTA) stained with ethidium bromide. Lane 1: 1 kb DNA ladder (New England Biolabs). Lane 2: 147-bp Widom 601 DNA nucleosome positioning sequence. Lane 3: nucleosome core particle after sucrose gradient purification.

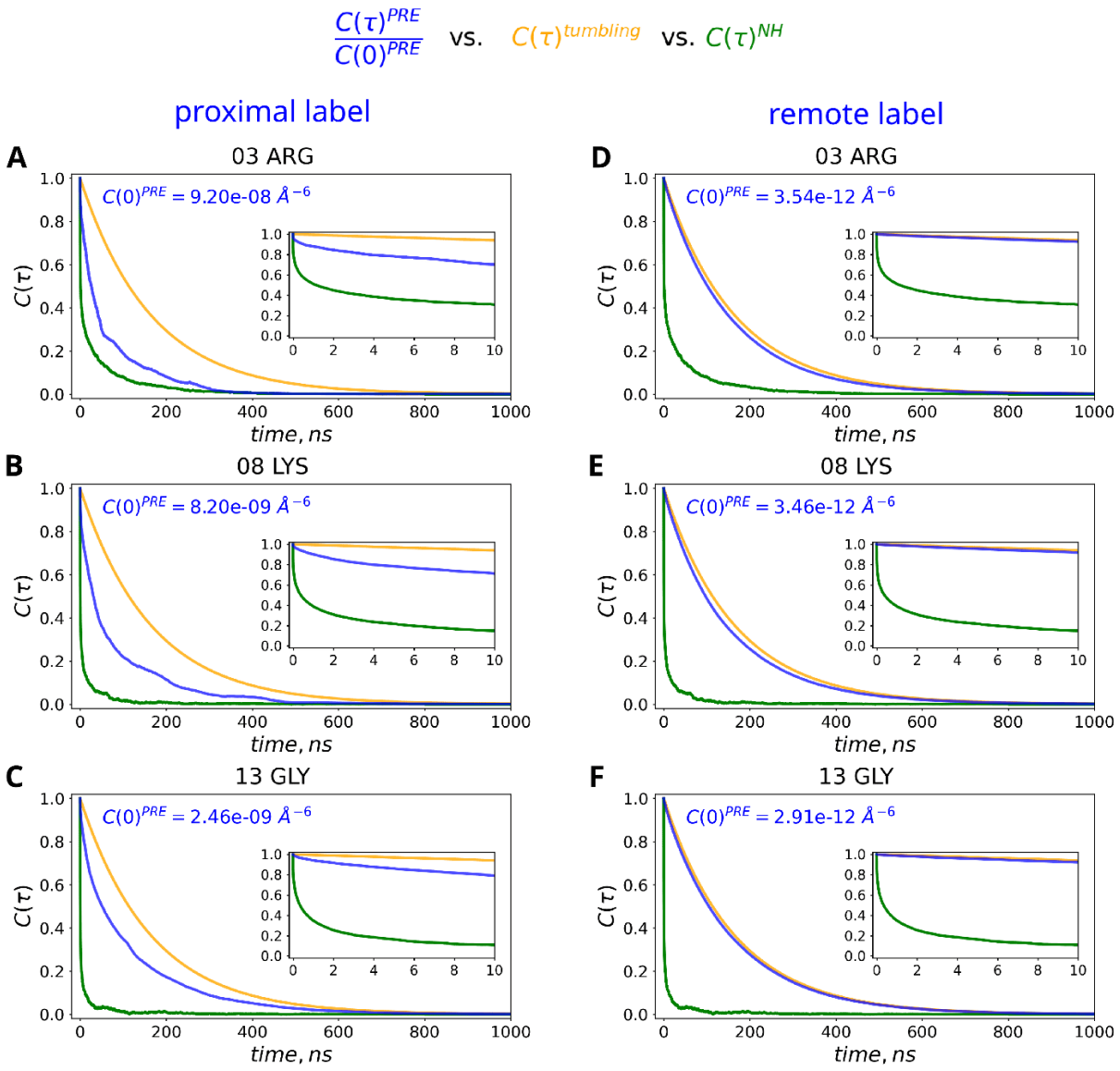


Figure 2.5 — Examples of the calculated PRE correlation functions $C(\tau)$ from 2-μs MD trajectory of NCP in TIP4P-D water (blue curves). The results are for the observation sites R3, K8 and G13 within the histone H4 tails in the NCP sample, which is spin-labeled at position L65C of histone H3; the contributions from proximal and remote labels are shown separately (see legends in the graph). For the purpose of plotting, the correlation functions are normalized by $C(0)$ (whose value is indicated in each panel). For comparison, we also show the correlation function of the NCP tumbling motion with the characteristic time τ_{rot} of 163 ns[50] (orange curves) and the correlation functions representative of $^{1}H^{N}$-^{15}N dipolar interaction for the three observed amide sites (green curves). Note that $^{1}H^{N}$-^{15}N correlation functions are essentially different from the PRE correlation functions. For example, consider the scenario where H4 tail is loosely anchored to nDNA through residue R3, so that the portion of the tail aa 3-15

effectively forms a large floppy loop. In this situation, the movement of the tail relative to the nitroxide label may occur only slowly, whereas the conformational dynamics of the loop-like segment can be fast. As illustrated by this example, nitrogen relaxation data cannot reliably predict the PRE-related dynamics. The correlation functions from OPC simulations (not shown) are similar to those from TIP4P-D simulations. However, they tend to decay considerably slower, consistent with our observation that H4 tails are less dynamic in the OPC trajectories.

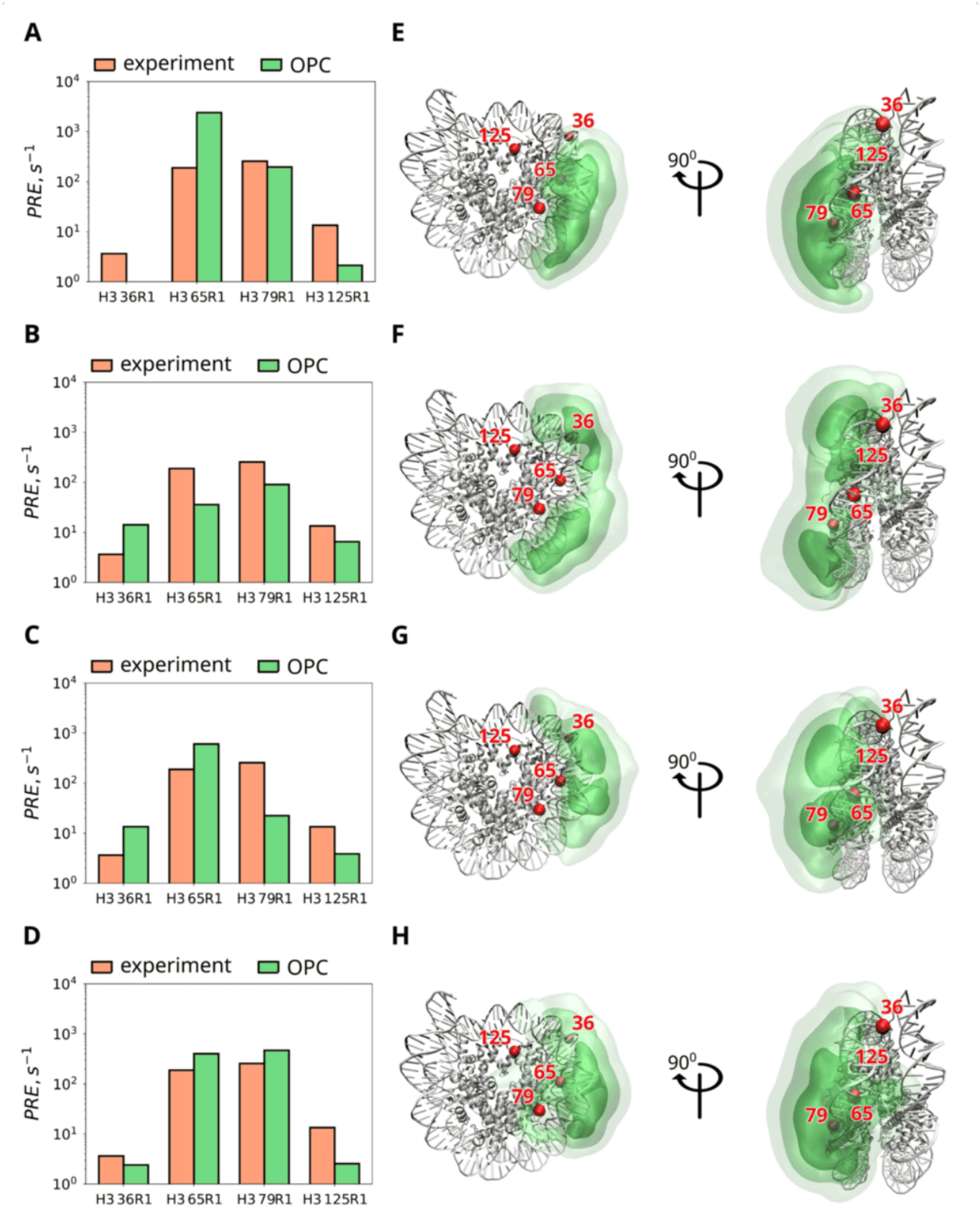


Figure 2.6 — (A-D) Experimental PRE rates averaged over H4 residues 3-15 (coral bars) and the corresponding MD predictions (green bars) based on four different subsets of OPC trajectories, originally labeled A-D (briefly summarized below, see ref. [102] for additional details). (E-H) The localization of H4 tails in four different subsets of OPC trajectories, A-D. The plotting conventions are the same as in Figure 2.2. The trajectories have been grouped according to the

original protocols used to build the initial NCP models [102]. Model A is based on the structure PDB:1KX5, where the original DNA sequence has been replaced with the extended DNA sequence from oncogene KRAS. The histone tails are from PDB:1KX5 and also transferred from PDB:1AOI. The missing residues at the end of the tails have been modeled by linearly extending the existing tail conformations. Model B is analogous to model A, except the histone tail conformations are from PDB:1KX5 and PDB:1EQZ. In model C, the partially resolved tails contained in PDB:1KX5 have been clipped and then rebuilt with a linear geometry following the backbone orientation of the last two residues at the truncation sites. The arising steric clashes are removed by means of the program Modeler [215]. Model D is analogous to model C, but with an additional requirement that the pairs of histone tails should be oriented symmetrically with respect to the dyad axis.

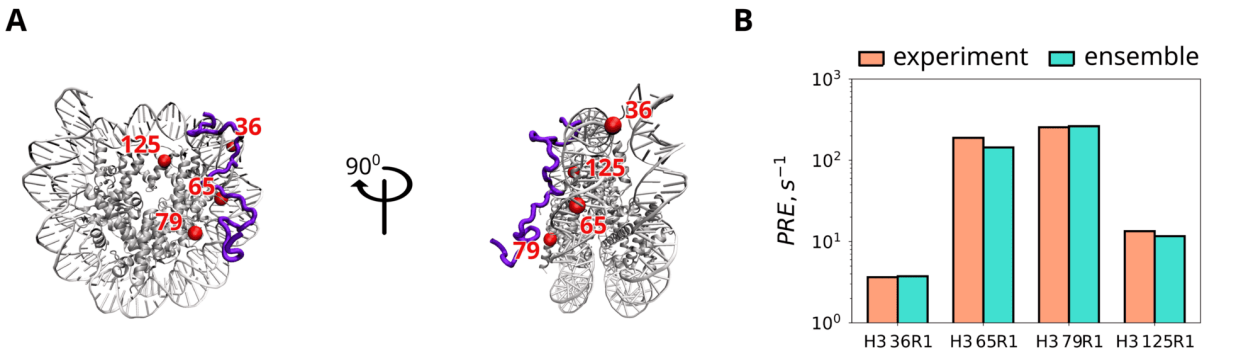


Figure 2.7 — (A) Two-member conformational ensemble of NCP predicting the PRE data in good agreement with the experiment. The ensemble is derived from the TIP4P-D trajectory as described in the section 2.3; the respective H4 tails are visualized in the plot (purple coils) along with the nitroxide label attachment sites (red spheres). The weights assigned to the two conformers are 93 % and 7 %. (B) Comparison of experimental and ensemble-based calculated PRE rates (coral and turquoise bars, respectively). The calculated PREs are obtained using the Solomon-Bloembergen expression for spectral density, equation (2.7), and averaged over H4 residues 3-15 and over the two conformers in the ensemble (taking into account their respective statistical weights). Although the two-member conformational ensemble successfully reproduces the experimental PRE data, it clearly fails to represent the conformational diversity of the H4 tail in the nucleosome core particle.

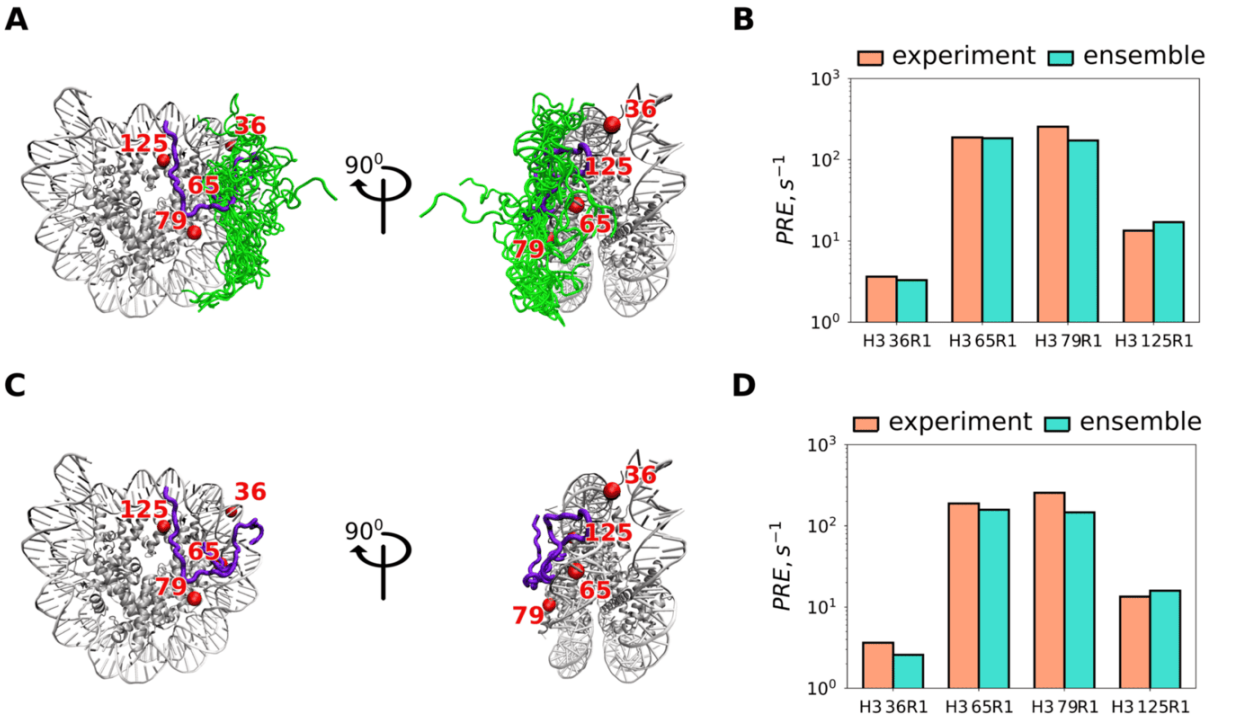


Figure 2.8 — (A) 42-member conformational ensemble of NCP predicting the PRE data in good agreement with the experiment. The ensemble is derived from the OPC trajectories as described in the section 2.3; the respective H4 tails are visualized in the plot (purple and green coils) along with the nitroxide label attachment sites (red spheres). The two conformers shown in purple are assigned the weights of 10 % each; they are responsible for the bulk of the PRE effect. The forty conformers shown in green are assigned the weight of 2 % each; they contribute only marginally to the calculated PRE rates. (B) Comparison of experimental and ensemble-based calculated PRE rates (coral and turquoise bars, respectively). The calculated PREs are obtained using the Solomon-Bloembergen expression for spectral density, equation (2.7), and averaged over H4 residues 3-15 and over the 42 conformers in the ensemble (taking into account their respective statistical weights). (C) Reduced conformational ensemble of NCP, containing two PRE-active conformers (purple coils, same as in panel (A)) with the assumed weights of 10 % each. Note that these two conformers are entirely different from the ones shown in Figure 2.7. (D) Comparison of experimental and ensemble-based calculated PRE rates (coral and turquoise bars, respectively) for the reduced conformational ensemble shown in panel (C). Although formally the 42-member ensemble successfully reproduces the experimental PRE data, it is in fact only minimally restrained and therefore cannot be viewed as a meaningful PRE based model of the H4 tail dynamics.

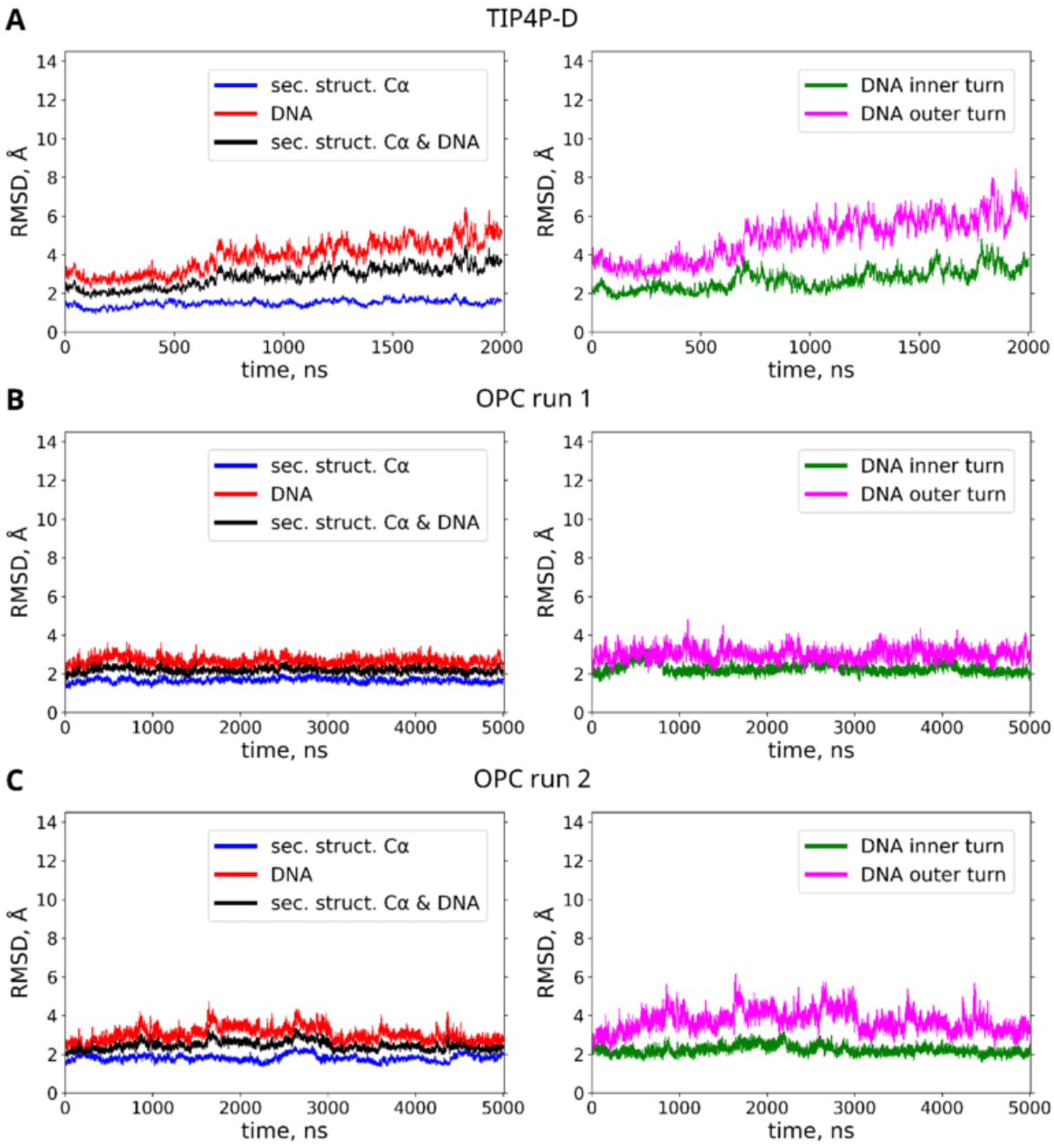


Figure 2.9 — (A) RMSD traces for 2-μs TIP4P-D trajectory of NCP. All frames in the trajectory have been superimposed onto the reference structure 3LZ0, using for this purpose only C^{α} atoms from the rigid histone scaffold (i.e. those C^{α} atoms that belong to the α-helical secondary structure). Subsequently, RMSD values have been calculated using different sets of atoms: (1) the same C^{α} atoms that have been used to overlay the frames (blue curve); (2) N1 and N9 atoms from nDNA nucleobases (red curve); (3) sets (1) and (2) combined (black curve); (4) N1 and N9 atoms from the inner turn of nDNA, nucleotides from -38 to 38 (green curve); (v) N1 and N9 atoms from the outer turn of nDNA, nucleotides from -72 to -39 and from 39 to 72 (magenta curve). The sampling step of 1 ns

has been used. (B, C) Analogous RMSD traces for two representative 5-µs OPC trajectories of NCP belonging to groups B and D in ref. [102], respectively. Note that the definition of the nDNA inner and outer turns is the same as before, disregarding the linker segments.

Table 7 — Experimental PREs for individual H4 tail amide sites in the four NCP samples containing paramagnetic spin labeled histone H3: 36R1, 65R1, 79R1 and 125R1. For each NCP sample, the PREs were extracted from residue-specific peak volumes in the ^{15}N-^{1}H^{N} HSQC NMR spectrum for the paramagnetic sample and a reference diamagnetic sample generated by addition of excess sodium ascorbate (c.f., Figure 2.1) as described above in the section 2.3. For the three strongly overlapped glycine resonances, G4, G6 and G13, the experimentally measured effective average PRE is reported. Note that for the H3 36R1 sample three residues had small negative PREs, which have been assigned a value of zero (with the values determined experimentally indicated in square brackets). Shown in the bottom two rows is the average PRE for all H4 tail residues for each sample, with the corresponding standard deviation.

	PRE(s^{-1})			
H4 residue	H3 36R1	H3 65R1	H3 79R1	H3 125R1
R3	2.5	194.1	255.3	0.3
G4/G6/G13	0.7	172.1	327.1	7.6
K5	11.3	105.0	216.4	14.9
G7	5.5	194.8	265.6	9.3
K8	4.9	165.3	350.5	15.4
G9	4.8	210.4	180.0	12.2
L10	10.8	229.0	239.0	13.2
G11	0.1	227.7	268.6	17.5
K12	0.0 [-1.6]	165.2	229.2	19.4
G14	0.0 [-2.2]	164.9	248.0	11.7
A15	0.0 [-2.4]	250.1	235.0	26.8
Average	3.7	189.0	255.9	13.5
Standard deviation	4.2	40.5	48.1	6.8

Chapter 3. Using NMR spectroscopy and MD modeling to study aromatic ring-flips in ubiquitin crystals

3.1 Introduction

Aromatic side chains play important roles in proteins. Often located in their hydrophobic cores, they are key to protein stability. Over-represented in protein-protein and protein-drug interfaces, aromatics play an important role in molecular recognition and binding [216], and are often prominent in the active sites of enzymes. The dynamics of aromatic side chains have been intensely studied research for more than four decades [217—220]. An important motivation for this interest is the realisation that aromatic side chains are rather bulky, and their motions, particularly rotations of the aromatic rings, require a significant void volume. Consequently, aromatic ring flips are thought to reveal coordinated movement of surrounding residues. Ring flips of His rings have also been studied in the context of functional mechanisms of enzymes and channels [221; 222].

Ring flips of Phe and Tyr, i.e. 180° rotations around the χ_2 dihedral angle, interconvert two indistinguishable states, and the exchange between these is, therefore, not observable by crystallographic methods. Nuclear magnetic resonance (NMR) spectroscopy can probe such motions in quite some detail, including the time scales of ring flips and the amplitudes and time scales of the ring-axis motions. Solution-state NMR provides insights into ring flips because the two symmetry-equivalent spin pairs (at the two H^{ε}-C^{ε} positions, or the two H^{δ}-C^{δ} positions, respectively, also denoted *ortho*-CH and *meta*-CH) are exposed to different conformational environments. Observing either two distinct sets of peaks, or a single time-average set of peaks, or possibly line broadening due to exchange, provides evidence for the time scale of flips. A growing arsenal of methods allows for quantification of such exchange processes [223—226]. Studying the pressure- or temperature-dependence of such parameters sheds light onto the transition state and created void volume involved in ring flips [227—229]. A recent study managed to stabilise a transition state of a ring flip which became then amenable to high-resolution structural investigation [230].

Nuclear spin relaxation methods [229; 231] provide another avenue to probe ring dynamics of proteins in solution.

Whether crystal packing has an influence on ring flips is poorly understood, largely because these are invisible to crystallography. Magic-angle spinning (MAS) NMR provides atom-resolved insight into protein assemblies, including crystals. It can, thus, shed light onto the impact of crystal packing on motions. MAS NMR has been used for studying dynamics of aromatic rings [220; 232—234] and protein dynamics more generally (see reviews, e.g. references [235—242]). We have recently applied a selective isotope-labeling strategy combined with sensitive proton-detected MAS NMR pulse sequences to quantitatively probe aromatic ring dynamics over a wide range of time scales [243]. The approach uses highly deuterated protein samples, in which ^{1}H-^{13}C spin pairs are introduced at either the C^{ζ} (*para*-CH), the C^{ε} (*meta*-CH) or the C^{δ} (*ortho*-CH) site. Together with MAS frequencies of 40-50 kHz or above, this strategy leads to sensitive high resolution ^{1}H-^{13}C correlation spectra. Moreover, given the simplicity of the spin system, with well-isolated ^{1}H-^{13}C spin pairs, it is straightforward to obtain quantitatively accurate dynamics data without any influence of scalar or dipolar couplings to remote spins. In the solid state, the arsenal of methods for probing dynamics is richer than in solution. In addition to chemical-shift based methods and relaxation measurements (also accessible in solution). MAS NMR allows quantifying dipolar couplings and provides insights into microsecond-millisecond dynamics from experiments that are sensitive to the MAS frequency and radiofrequency (RF) fields, such as NEar-Rotary resonance Relaxation Dispersion (NERRD) experiments [244; 245] (see below). Measurements of dipolar couplings are very useful to learn about motional amplitudes: motion leads to averaging of dipolar couplings, and the averaged dipolar-coupling tensor reflects the conformational space that the inter-atomic vector samples. The methods we employ here allow to even see anisotropy of the underlying motion, that is caused e.g. by two-site ring flips [243; 246]. Additionally, MAS NMR relaxation measurements probe dynamics over a broad range of time scales without inherent «blind spots» [236—238]. Although MAS NMR relaxation experiments are differently sensitive to different time scales [247; 248], in principle any time scale can be probed. In particular, NERRD experiments allow probing whether motions occur on microseconds (μs) or rather nanosecond (ns) time scales [248].

Here, we use MAS NMR together with highly deuterated samples with specific ^{1}H-^{13}C spin pairs to investigate phenylalanine ring dynamics in the ubiquitin protein, crystallized in three different crystal forms, herein called MPD-ub, cubic-PEG-ub and rod-PEG-ub. These names refer to the crystallisation agent, methyl pentanediol or polyethylene glycol, and the crystal shape. The arrangement of the molecules, in particular with respect to the Phe side chains, is displayed in Figure 3.1.

These three crystals differ in the number of molecules inside the unit cell, the relative orientation of molecules to each other and the solvent content in the crystal, ranging from 58 % water content in cubic-PEG-ub to 49 % (MPD-ub) to 40 % (rod-PEG-ub). Previous studies of the backbone dynamics of these crystals have revealed that the backbone has similar (sub-µs) motions, but that there is different degree of overall motion: the molecules of the least densely packed cubic-PEG-ub crystal undergo overall rocking motion with an amplitude of several degrees on a µs time scale [244; 249]. While both Phe residues are positioned more or less on the protein's surface, rather than in the hydrophobic core, the packing of the aromatic side chains differ. Using the highly accurate solution-NMR structure of ubiquitin (PDB ID: 1D3Z), we find that the average solvent accessible surface area (SASA) of Phe4 in this multi-conformer structure is 59 Å^2 (33 %), while for Phe45 it is 42 Å^2 (23 %). However, the degree of solvent exposure changes depending on intermolecular arrangements in the different crystal forms (Figure 3.1 (D-I)), which is also reflected in the number of contacts that each Phe side chain makes with the neighboring ubiquitin molecules (Figure 3.1 (J)).

MAS NMR combined with MD simulations, presented herein, shed light onto the effects of crystal packing on Phe ring dynamics. The entirety of spectroscopic data suggest that the Phe45 signals are broadened beyond detection in all crystals due to slow (µs) ring flips of Phe45. This finding is backed up by MD simulations. The peak positions of Phe4 are remarkably conserved in the three different crystal forms, despite different buffer composition and pH conditions and intermolecular packing. However, in one crystal form, differences in intermolecular packing between the two chains are reflected by different ^{1}H chemical shifts and spin relaxation parameters. Overall, our study reveals that the impact of the crystal packing on aromatic ring dynamics are detectably impacted by the crystal packing, but the effect is small compared

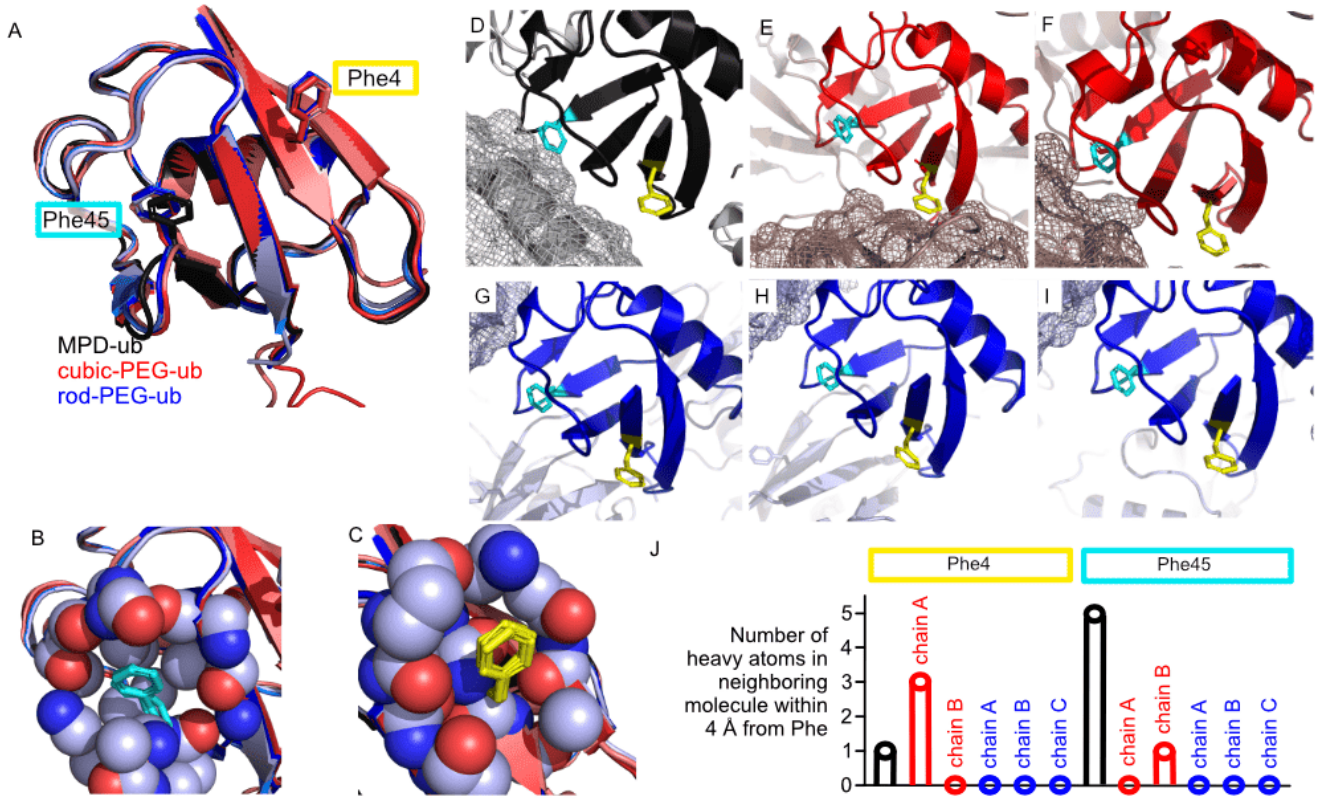


Figure 3.1 — Structures of ubiquitin in three different crystal forms, denoted herein as MPD-ub (black), cubic-PEG-ub (red) and rod-PEG-ub (blue). The three crystal forms correspond to PDB entries 3ONS, 3N30 and 3EHV, respectively. The number of non-equivalent molecules in the asymmetric unit are: 1 (MPD-ub), 2 (cubic-PEG-ub) and 3 (rod-PEG-ub). (A) Overlay of the backbones of the six (1+2+3) chains from the three crystal forms. Panels (B) and (C) zoom onto Phe45 and Phe4, respectively. Spheres denote atoms within 5 Å around the aromatic side chain (shown here for one of the chains of rod-PEG-ub; dark blue: nitrogen, red: oxygen, light blue: carbon). (D-I) Crystal packing for MPD-ub (D), cubic-PEG-ub, chains A (E) and B (F), and rod-PEG-ub, chains A (G), B (H) and C (I). Neighboring molecules in direct contact with the molecule in the center are shown as mesh. (J) The number of heavy atoms located in a neighboring ubiquitin chain within a radius of 4 Å around any atom of Phe4 (left) or Phe45 (right).

to the intramolecular determinants of ring flipping. Interestingly, the strong difference of ring-dynamics time scale of Phe4 and Phe45 that we detect from NMR measurements and MD simulations is not directly related to the

solvent-accessible surface area, nor the rotameric state, which points to other intramolecular determinants of ring flips.

3.2 Results

3.2.1 Specifically isotope-labeled Phe samples of three crystal forms of ubiquitin

In order to obtain sensitive and high-resolution proton-detected MAS NMR spectra, combining high levels of sample deuteration with fast magic-angle spinning is an established method [250; 251]. For exchangeable sites, in particular the amides of the backbone, perdeuteration followed by back-exchange in $^{1}H_2O$ buffer is a straightforward method. For detecting side chain atoms, one can either use the "imperfection"of deuteration, i.e. the residual ^{1}H content in deuterated samples [252; 253], or specific labeling with precursors in which a chosen type of moiety is protonated almost completely. The latter approach is commonly used for methyl-directed NMR, particularly in solution NMR [254; 255], and also in MAS NMR [246; 256]. Introduction of such isolated protons in other sites allows to obtain highest resolution for other side chains; generally, such specific labeling approaches can clearly achieve better line widths than those obtainable from fully protonated samples at the highest available MAS frequencies to date (100 kHz) [243].

We have recombinantly expressed ubiquitin in which all non-exchangeable sites are deuterated, and individual ^{1}H-^{13}C spin pairs are incorporated at the two H^{ε}-C^{ε} positions of Phe residues. We denote this labeling here as u-[^{2}H,^{15}N],Phe(ε1,ε2)-^{1}H,^{13}C. The incorporation of the specific label was achieved by adding a properly labeled ketoacid precursor molecule (35 mg per liter of culture), displayed in Figure 3.2 (A), to the bacterial culture prior to induction [257], as described in the section 3.3.

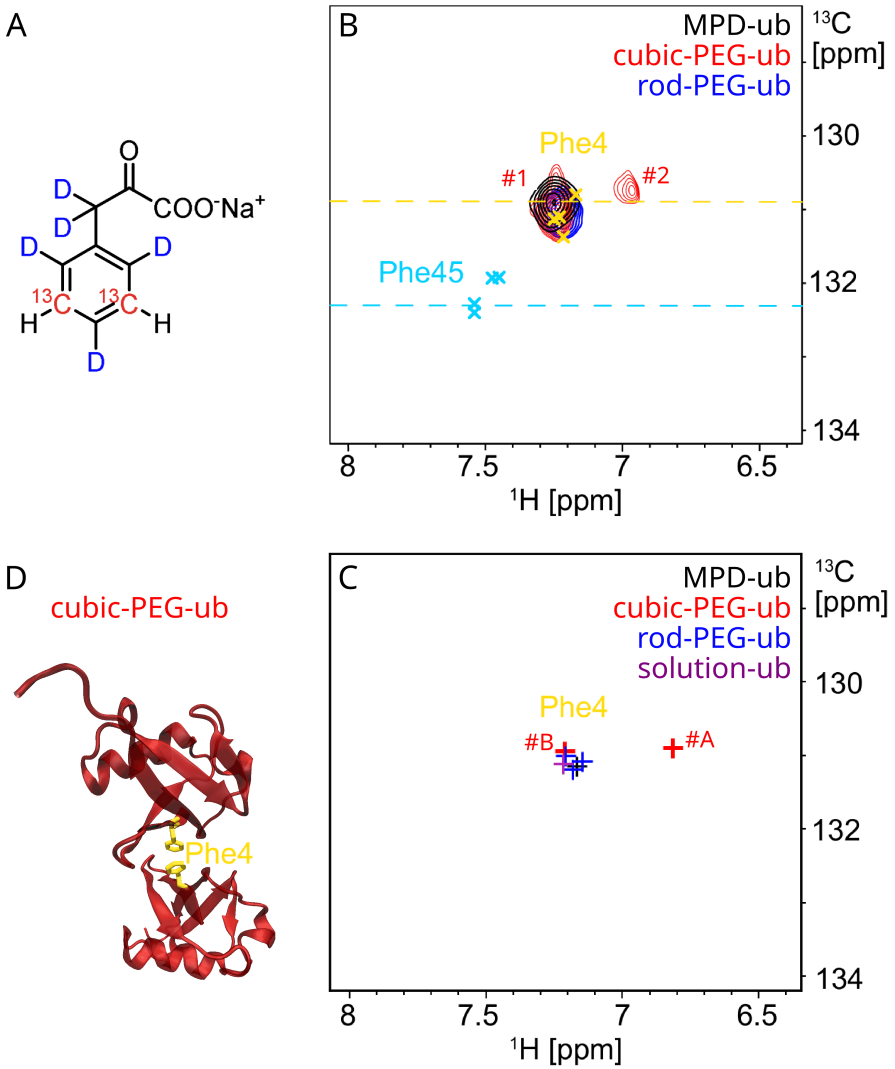


Figure 3.2 — MAS NMR of specifically Phe-labeled ubiquitin. (A) Ketoacid precursor used for the labeling of Phe with two ^{1}H-^{13}C spin pairs at the two ε positions. (B) Overlay of the cross-polarization (CP) based $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ correlation spectra of the three different ubiquitin crystal samples. The spectra were obtained at 50 kHz MAS frequency at a sample temperature of ca. 28 °C with a pulse sequence based on ^{1}H-^{13}C out-and-back cross-polarisation steps and ^{1}H detection [243]. MPD-ub and rod-PEG-ub spectra feature a single observable peak, while cubic-PEG-ub displays two peaks (labeled #1 and #2 in the plot). The assignments of Phe4 and Phe45 in various solution-state data sets (BMRB 6457, 5387, 16228 and 27356) are indicated by crosses, and the solid-state NMR assignment (only for ^{13}C; BMRB 25123) is indicated by a horizontal line, color-coded for Phe4 (gold) and Phe45 (cyan). (C) Chemical-shift predictions for $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ spin pairs in Phe4 using the program SHIFTX2. Tthe peaks are color-coded as indicated in the legend. Of interest, chain A in cubic-PEG-ub shows a distinctive proton chemical shift, significantly different

from that of chain B. This effect can be attributed to ring current shift arising from the stacking of the two Phe4 rings from the proximal chains A in the crystal lattice (illustrated in panel (D)).

Figure 3.2 B shows the aromatic region of the ^{1}H-^{13}C spectra of ubiquitin in the three different crystal forms. Overlaid with these spectra are previous assignments of the $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ sites in solution [258—261], and of the $^{13}C^{\varepsilon}$ in a carbon-detected MAS NMR study [262]. Even though these previous data sets have been collected under a diverse set of conditions, including in-cell NMR and a sample in reverse micelles, the reported peak positions are remarkably similar.

3.2.2 Phe4 rings undergo sub-millisecond ring flips

In obtained spectra, only the signal that corresponds to Phe4 is visible, while the one of Phe45 is absent. We discuss the question of why the signal of Phe45 is not observed further below. We observe a single peak for Phe4 in both MPD-ub and rod-PEG-ub spectra. For MPD-ub, in which all chains are identical, this is expected. However, this outcome is less obvious for rod-PEG-ub, where the crystal contains three non-equivalent molecules in the unit cell and many backbone amide sites show three distinct signals [249]. Closer inspection of the three non-equivalent Phe4 appearances in the crystal (see Figure 3.1 (G-I)) shows that the Phe4 side chains do not have any contact to other molecules. Thus, their environment is very similar in the three non-equivalent molecules, which explains the observation of a single signal. In cubic-PEG-ub we observe two signals: one overlaps with the position found in the other crystal forms and in solution (denoted as (#1) in Figure 3.2), and the other one shifted by ca. 0.25 ppm upfield in proton dimension (#2). These two peaks mirror the differences in the environment of Phe4 in the two non-equivalent molecules in the crystal (Figure 3.2 (E, F, J)). We tentatively assign peak (#1), which is very close to the solution-NMR position, to chain B, because in chain B Phe4 does not form any intermolecular contacts, and peak (#2) to chain A, which is engaged in contacts to a neighboring chain.

To confirm this view, we have performed SHIFTX2 predictions [263] of the $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ correlations in solution and in the three crystal forms. Chemical-shift prediction programs are often challenged to predict proton frequencies with accuracy, particularly those in side chains, for which fewer data are available. However, the predictions are able to reproduce the effects observed for Phe4 rather well. In particular, the prediction finds a 0.39 ppm upfield shift of the signal corresponding to chain A of cubic-PEG-ub, compared to chain B. This effect is due to intermolecular stacking of Phe4 rings in the two neighboring chains A (Figure 3.2 (D)). When the second chain is removed in the SHIFTX2 calculation, the two Phe4 signals land essentially on top of each other. The effects of ring-currents are well understood [264; 265], and we assume that the effects of intermolecular ring stacking are well recapitulated in the predicted Phe4 shifts of cubic-PEG-ub.

On the other hand, the limited accuracy of the structure-based chemical shift predictors is apparent in the results for Phe45. While SHIFTX2 correctly predicts that Phe45 shifts are similar among the different crystal forms and in solution, their absolute values do not agree very well with the experiment, see Figure 3.7.

The observation of a single cross-peak for Phe4, which is labeled at the two εsites, suggests that the isotropic chemical shift of the two positions is averaged by sub-millisecond ring flips. To gain a more direct insight into ring flips, MAS NMR can probe the dipolar-coupling averaging. The ^{1}H-^{13}C dipolar-coupling tensor is averaged by motions faster than ca. 10-100 µs (Figure 28 in ref. [237]). For the case of ring flips the tensor anisotropy is reduced to theoretically 62.5 % (order parameter S=0.625); moreover, the dipolar-coupling tensor, which is uniaxial in the rigid-limit case (i.e. axially symmetric), becomes biaxial. We use here the term biaxial; this property is also termed tensor asymmetry in the literature, with an asymmetry parameter, η, where η=0 denotes an axially symmetric tensor. As ”biaxial”more precisely reflects the shape of the tensor, and in line with the use in other fields of physics, [266] we use the term biaxiality herein. One can show that the ring flips not only reduce the anisotropy of the dipolar-coupling tensor to S=0.625, but also increase the biaxiality parameter to a value of η= 0.6 [243]. An adapted version of the Rotational Echo DOuble Resonance (REDOR) experiment [267] allows determining order parameters and tensor biaxiality parameters [246]. Figure 3.3

(A) shows the REDOR curves for Phe4 in the three crystal forms. In all cases, the obtained tensor parameters are in agreement with the ring flip model, i.e. the order parameter matches (within error bars) the expected value of 0.625, and the confidence interval of the biaxiality parameter encompasses the expected value of 0.6, see Figure 3.3 (B). In some crystal forms the order parameters and the biaxiality are somewhat lower than in others, possibly due to additional small-amplitude motions, but it is clear that within error bar the ring flips alone can account for the experimental observations.

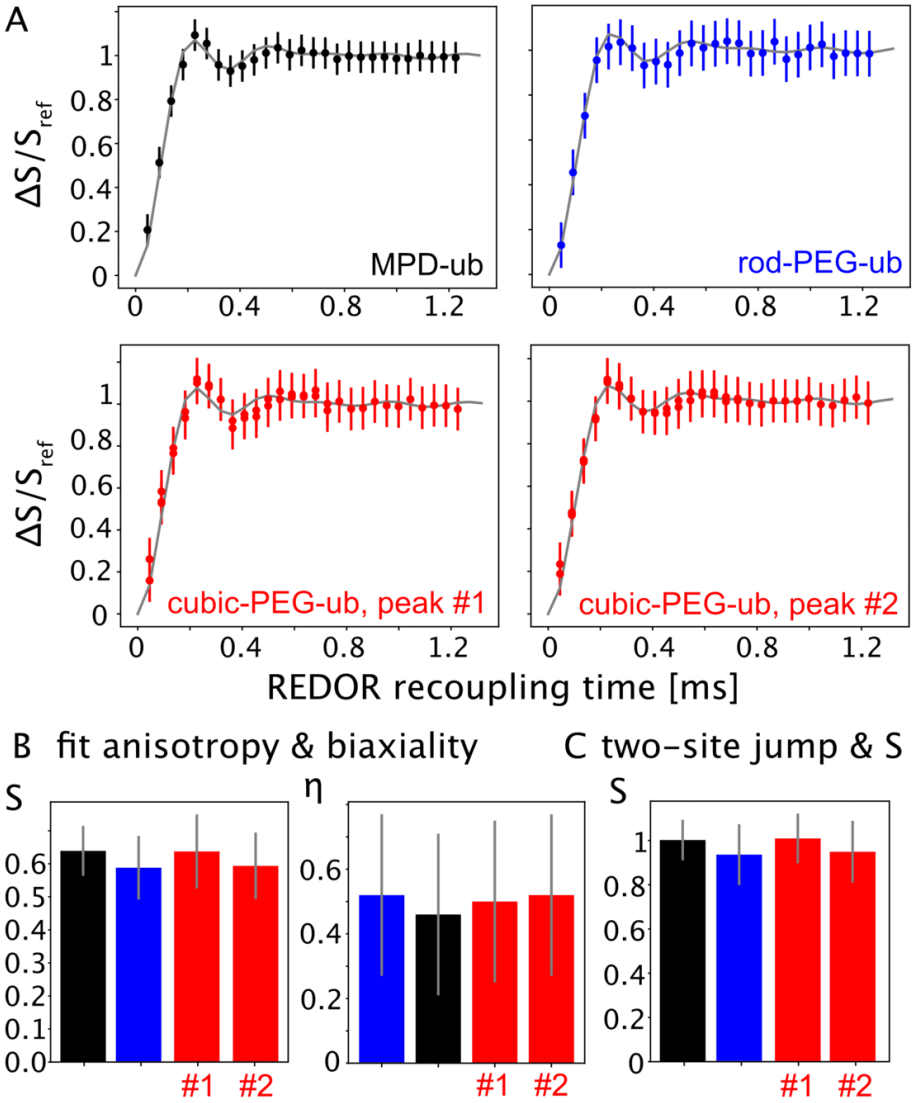


Figure 3.3 — ^{1}H-^{13}C dipolar-coupling tensor measurements for the ε site in Phe4. (A) REDOR recoupling curves for the different crystals. (B) Fitted tensor parameters in a fit that does a grid search for the best order parameter S (i.e. the tensor anisotropy) and tensor biaxiality parameter η. (C) Fitted order parameter from a grid search against a grid of simulations that assume explicit two-site jumps (120°) and a variable tensor anisotropy δ_D. The resulting best-fit order parameter (calculated, as usual, as $S=\delta_{D,fitted}/\delta_{D,rigid}$) is close to 1, reflecting that

the tensor is controlled by the two-site jumps, with only very small additional motional averaging.

3.2.3 Phe4 ring flips occur on a 10-20 ns time scale

Spin relaxation rate constants are sensitive to amplitudes and time scales of motion. We have measured the ^{13}C longitudinal (R_1) and rotating-frame ($R_{1\rho}$) relaxation, as well as ^{1}H $R_{1\rho}$, and used them to determine the ring-flip rate constants. To this end, we compared the experimentally measured ^{13}C relaxation rate constants (Figure 3.4 (A, B) and Table 8) to calculated rate constants that result from ring flips (Figure 3.4 (D)).

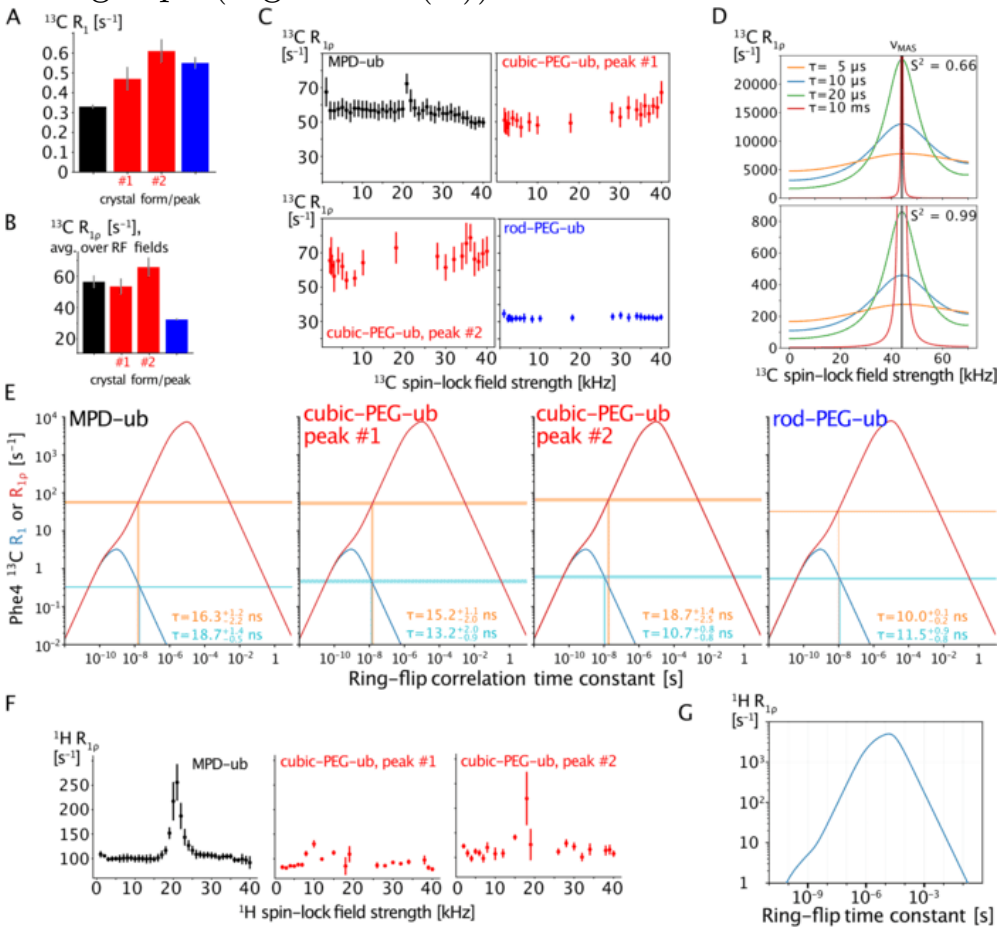


Figure 3.4 — MAS NMR dynamics data for Phe4 signals in the three different crystal forms. All data have been collected at 44.053 kHz MAS frequency. In all cases, the colors black, red and blue refer to data from the three different crystal forms. (A) ^{13}C longitudinal (R_1) and (B) rotating-frame ($R_{1\rho}$) relaxation

rate constants. (C) Measurements of ^{13}C $R_{1\rho}$ as a function of the spin-lock field strength (relaxation dispersion). The profiles do not show a marked increase when the RF field strength approaches the condition $\nu_{RF}=\nu_{MAS}$, which one would expect if the flip motion was on the µs time scale. (D) Calculated ^{13}C NERRD profiles (see section 3.5.1) for different correlation times and an order parameter corresponding to ring flips (S=0.66, upper panel) or smaller amplitude motion (S=0.99, lower panel). The rotary resonance condition is indicated with a black vertical line at 44.053 kHz. (E) Determination of the ring-flip correlation times for Phe4 from ^{13}C relaxation rate constants. The Λ-shaped profiles show relaxation rate constants calculated for ring flips as a function of the time scale of these flips; orange: $R_{1\rho}$; blue: R_1. Horizontal lines indicate the experimentally measured rate constants and vertical dashed lines show where the experimental data intercept the calculated curve (on the "fast"branch of the curve). (The numerical values of the experimental relaxation rates are summarized in Table 8. Note the remarkable agreement of the flip time constants from $R_{1\rho}$ and R_1. (F) ^{1}H $R_{1\rho}$ relaxation-dispersion profiles. There is a rise of $R_{1\rho}$ at $\nu_{RF}=\frac{1}{2}\nu_{MAS}$ ("HORROR-condition"[268]), which is due to the recoupling of the ^{1}H-^{1}H dipolar coupling, visible in particular for the MPD-ub crystal data. The width of this feature is limited to only a few kHz around the halved spinning frequency. (G) Calculated ^{1}H $R_{1\rho}$ rate constants (see section 3.5.1 for details).

In these calculations, we fixed the order parameter of the ^{1}H-^{13}C moiety to the one expected for ring flips, and varied the corresponding ring-flip correlation time. The calculated relaxation rate constants for a correlation time in the 10-20 ns range match the experimental ones for all crystal forms and for the two sites of cubic-PEG-ub. The correlation times of the ring flips determined from $R_{1\rho}$ and R_1 agree very well with each other, supporting that a single dominant motion, ring flips, accounts for the bulk of spin relaxation. We have considered the possibility that additional motions, other than ring flips, contribute to the observed relaxation rate constants. To explore this possibility, we have extended our model by including additional motional modes as found in our MD simulations (see Figure 3.8 and section 3.5.2). Re-analyzing $^{13}C^{\varepsilon}$ R_1 and $R_{1\rho}$ using this extended model shows that additional motions have only minimal influence on the extracted ring-flip rates, see Figure 3.9.

Even though these data clearly point to nanosecond flips, we also probed whether Phe4 undergoes µs motions, possibly of very small amplitude, using ^{13}C $R_{1\rho}$ NERRD experiments [244; 245]. In these experiments, the $R_{1\rho}$ rate constant is probed as a function of the spin-lock RF field strength. In the presence of µs motion, $R_{1\rho}$ increases sharply when the RF field strength approaches the MAS frequency. A ^{1}H $R_{1\rho}$ NERRD version has been proposed, too, and, although less straightforward to quantify, is another way to detect µs motions [269; 270]. Figure 3.4 (C, F) show the ^{13}C and ^{1}H NERRD data, respectively. These profiles are flat in most cases, suggesting that there is no significant µs motion. The observed rise in the ^{1}H NERRD profiles at the condition $\nu_{RF}=\frac{1}{2}\nu_{MAS}$, is due to the recoupling of the homonuclear dipolar coupling at the HORROR condition [268]; however, it extends over only a narrow range of RF field strengths, much less than what is expected if the motion occurred on a µs time scale [270]. It is noteworthy that the experimentally observed ^{1}H $R_{1\rho}$ rate constants for Phe4, ca. 50-120 s^{-1}, are higher than the expected ones for ring flips occurring on the ca. 10-20 ns time scale (ca. 30-50 s^{-1}; see Figure 3.4 (G)). This suggests that dipolar dephasing [271; 272] appears to be responsible for more than half of the expected decay rate constant.

The calculations illustrated in Figure 3.4 also offer a plausible explanation why the signal of Phe45 is unobserved. If ring flips for Phe45 are 1-2 orders of magnitude slower than for Phe4 (MD suggests a factor of ca. 50; see below), then the relaxation time constants of ^{1}H and ^{13}C at this site are expected to be of the order of 1 ms or less (and additional dipolar dephasing would shorten the ^{1}H life time even more). Such fast relaxation would not lead to coherence transfer through the experiment, and broaden signals beyond detection. Hence, Phe45 magnetization would decay rapidly during the pulse sequence and detection, obliterating the spectral signal.

Of note, in cubic-PEG-ub peak #2 (assigned to chain A) has higher ^{1}H and ^{13}C $R_{1\rho}$ rate constants than peak #1. This is likely the basis why the peak intensity of peak #2 is lower. Given that the ring flip rates are very similar, the origin of this faster relaxation is not entirely clear; a likely reason could be the closer proximity of protons from the neighboring molecule in the crystal, see Figure 3.1 (J).

Lastly, it is noteworthy that the molecules in the cubic-PEG-ub crystals undergo rocking motion within the crystal, while for the more densely packed

MPD-ub and rod-PEG-ub crystals this was not found [244; 249]. This rocking effect, occurring on a time scale of tens of µs, was detected by non-flat ^{15}N NERRD profiles. In this sense, it is interesting that the ^{13}C NERRD profiles of cubic-PEG-ub (but not of the other two crystal forms) show a slight (ca. 10 s^{-1}) increase toward $\nu_{RF}=\nu_{MAS}$ = 44 kHz. We ascribe this small NERRD effect to the overall rocking motion sensed by the side chain; the effect is quantitatively somewhat smaller than expected from the rocking parameters obtained for the backbone, which might reflect that the side chain is to some degree decoupled from the motion that the backbone senses. It is also possible that the contribution from rocking motion is more difficult to see for these aromatic ^{1}H-^{13}C sites than for backbone amides, because the absolute rate constants are ca. 10-fold larger than those of amide ^{15}N.

Overall, NMR data show that in all crystal forms the flips of Phe4 are similar, occurring on a 10-20 ns time scale. It appears that the flips of Phe45 occur on a slower time scale, which prevents its detection by NMR. The MD evidence for that is reviewed in the next section.

3.2.4 MD simulations provide insight into ring-flip dynamics

Molecular dynamics simulations provide a useful additional view on ring flips. We have analyzed microsecond-long trajectories of the explicit crystal lattices, as well as of ubiquitin in solution, in order to understand the observed differences in ring-flip rates, in particular between Phe4 and Phe45; the latter is experimentally not observed. Figure 3.5 shows the time traces of the χ_1 and χ_2 dihedral angles of the two Phe sidechains in the different crystal lattices and in solution. As the simulated crystal lattices comprise 24 or 48 molecules, and the simulation extends over 2 µs, the trajectories effectively cover many microseconds.

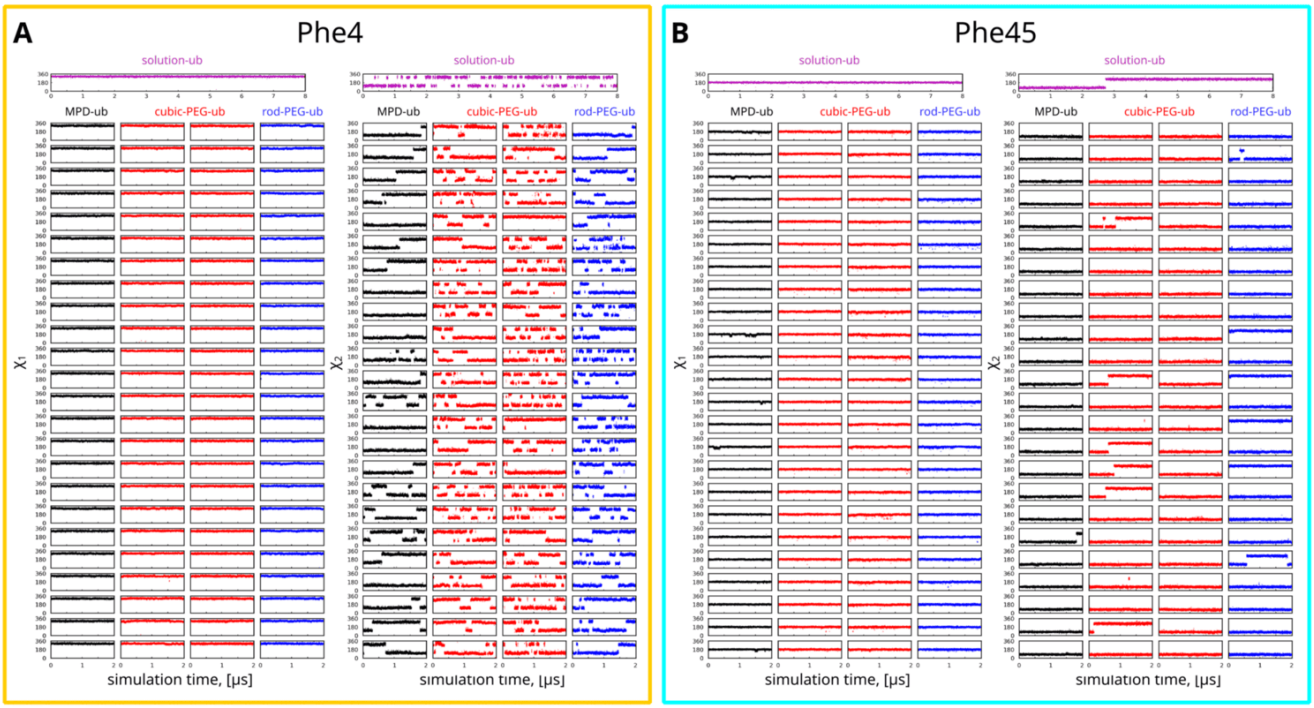


Figure 3.5 — Time traces of side-chain torsional angles χ_1 and χ_2 for (A) Phe4 and (B) Phe45 from the three simulated ubiquitin crystals and one simulation of ubiquitin in solution. To better visualize rotameric jumps, we use the angle range [0–360°] instead of the conventional choice [-180–180°]. The color coding is the same as in the previous figures: black (MPD-ub), red (cubic-PEG-ub), blue (rod-PEG-ub) and magenta (solution form). The cubic-PEG-ub simulation cell contains 48 ubiquitin molecules equally divided between chain A (molecules 1–24) and chain B (molecules 25–48). The rod-PEG-ub simulation cell contains 24 ubiqutin molecules equally divided between chain A (molecules 1–8), chain B (molecules 9–16) and chain C (molecules 17–24). The details on crystal simulations are provided in the section 3.3.

The χ_1 angle, which reorients the ring axis C^{β}-C^{γ}, does not change in any of the simulated systems. For χ_2, which represents rotations of the ring (flips and small-amplitude motions within the potential energy wells), the situation is more interesting and more diverse. Phe4 undergoes ring flips in all cases. In cubic-PEG-ub and rod-PEG-ub, these flips occur multiple times along the 2 μs long trajectory. In MPD-ub, which experimentally behaves very similarly, the flips occur less frequently than in the PEG-ub crystals.The aggregate estimate of the characteristic time for Phe4 ring flips using all crystal trajectories and solution trajectory is τ_{div}=204 ns (see Table 9 for the flip rate constants of the

individual simulations; τ_{div} is the total aggregated simulation time divided by the number of observed flips). This is an order of magnitude longer than the value of 10-20 ns estimated from the experimental relaxation data. Such difference translates into excess barrier height of ca. 1.5 kcal/mol, which is common even for state-of-the-art force fields [273].

In all crystals as well as in solution, the ring flips of Phe45 are much less frequent than those of Phe4. Specifically, the MD data indicate that there is a 47-fold reduction of the rate constant. While this value is to be considered a rough estimate, due to the insufficient sampling of the flip events in Phe45, the MD data show unambiguously a slow-down compared to Phe4. This kind of slow-down shifts the process into a range that is expected to cause a dramatic broadening of Phe45 signals, preventing its detection in MAS NMR spectra, as further discussed below.

3.3 Methods

3.3.1 Protein expression and purification

Perdeuterated ubiquitin with specific labels at the phenylalanine εpositions (^{1}H-^{13}C) was prepared by bacterial overexpression as follows. *Escherichia coli* BL21(DE3) cells were transformed with a pET21b plasmid carrying the human Ubiquitin gene. Transformants were progressively adapted in four stages over 48 h to M9/D_2O media containing 1 g/L $^{15}ND_4Cl$, 2 g/L D-glucose-d_7 as the sole nitrogen and carbon sources. In the final culture, the bacteria were grown at 37 °C. When the optical density at 600 nm (OD_{600}) reached 0.65, 35 mg of the ketoacid precursor (shown in Figure 3.2 (A)) per liter of culture were added. One hour later, while shaking at 37 °C the OD_{600} reached 0.95, whereupon protein expression was induced by addition of IPTG to a final concentration of 1 mM. Induction was performed for 3 h at 37 °C. At the end, the final OD_{600} reading was 2.2.

After induction, the cells were resuspended in 20 mL of 50 mM Tris-HCl pH 8 buffer containing 2 µg/mL leupeptine and 2 µg/mL pepstatine, and lysed

by sonication. The lysate was centrifuged for 30 min at 46,000 g using JA25-50 Beckman rotor, then the supernantant was dialyzed against two times 300 mL of 50 mM Tris-HCl pH 8 buffer. After dialysis the sample was centrifuged for 30 min at 46,000 g and loaded on a 40 mL Q-Sepharose column. Ubiquitin was recovered in the flow through fractions, which were subsequently concentrated and injected on a HiLoad 16/60 Superdex 75 column equilibrated with 1 column volume of 50 mM Tris-HCl pH 8 buffer. The protein was dialyzed against Milli-Q Ultra pure water until the buffer was completely removed. Then the protein was freeze-dried for 24 h.

3.3.2 Protein crystallization

The three different crystal forms, which were also used in our previous backbone-dynamics study although with different isotope labeling [249], were obtained by sitting-drop crystallisation with buffer conditions described below. In all cases, the crystals were obtained using sitting-drop crystallisation plates (Hampton research Cryschem plate, catalog number HR3-159) with a 40 µL sitting drop and 450 µL of reservoir buffer.

For preparing MPD-ub crystals ubiquitin was dissolved in buffer A (20 mM ammonium acetate, pH 4.3) at a concentration of 20 mg/mL. Buffer B (50 mM citrate, pH 4.1) was prepared and mixed with methyl pentanediol (MPD) at a volume ratio of 40:60 (buffer B : MPD), and 450 µL of this mix was placed in the reservoir of the wells. In the sitting drop, 37 µL of the ubiquitin / buffer A solution was mixed with 10 µL of the buffer B / MPD solution. The plate was covered with CrystalClear adhesive tape and kept at 4 °C. After ca. 1-2 weeks, needle-shaped ("sea-urchin like") crystals appeared.

For preparing cubic-PEG-ub crystals, the reservoir contained 450 µL of buffer C (100 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 6.3, 20 % (weight) PEG 3350 and 100 mM zinc acetate). The protein solution (20 mg/mL of ubiquitin) in buffer D (20 mM ammonium acetate, pH 4.8) was mixed with buffer C at a 1:1 ratio, and 45 µL thereof were placed in the sitting-drop holder. Cubic-shape crystals were obtained within 1 week at ca. 23 °C.

For preparing rod-PEG-ub crystals, the reservoir buffer contained buffer E (50mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0, 25 % PEG 1500 and 25 mM zinc acetate). The protein was dissolved in buffer D at 20 mg/mL, and mixed with reservoir buffer E (1:1), akin to the cubic-PEG-ub procedure.

Protein crystals were transferred to a custom-built ultra-centrifuge tool (essentially a funnel placed on top of a 1.3 mm Bruker NMR rotor, with dimensions that fit the buckets of a Beckman SW32 rotor). The crystals of each kind (ca. 3 mg protein) were centrifuged into their individual rotors, and the rotor caps were glued with two-component epoxy glue to avoid loss of water.

3.3.3 NMR experiment

All experiments were performed on a Bruker Avance III spectrometer operating at 600 MHz ^{1}H Larmor frequency (14.1 T) with a Bruker 1.3 mm probe where the main coil was tuned to ^{1}H, ^{13}C and ^{15}N frequencies, and an auxiliary coil to ^{2}H frequency. The MAS frequency was set to 40-50 kHz (specified in the figure panels and below) and maintained constant to within less than 10 Hz. The effective sample temperature was ca. 28 °C. The temperature was determined from a non-temperature sensitive signal of MPD and the bulk water line, using the equation T[°C] = 455 - 90·δ_{H2O}, where δ_{H2O} is the shift of the bulk water line in parts-per-million (ppm).

The pulse sequences for the proton-detected ^{1}H-^{13}C correlation experiments (hCH) have been presented in Figure 3.8 of ref. [243]. They include ^{1}H excitation, cross-polarisation to ^{13}C for chemical-shift editing (with ca. 10 kHz ^{1}H WALTZ-16 decoupling), flip-back of ^{13}C coherence to $^{13}C_z$ for ca. 40 ms water suppression by a train of ^{1}H pulses with 18 kHz field-strength amplitude and a duration of 820 µs, alternating in phase (±15°). The indirect ^{13}C dimension was typically sampled for ca. 12-15 ms, using a spectral width of 15 ppm (2250 Hz); the ^{1}H dimension was sampled for ca. 50 ms.

The cross-polarisation steps (H to C and C to H) used RF fields of ca. 85 kHz and 35 kHz at 50 kHz MAS frequency, and a duration of 400 µs, with a ramp (90 % to 100 %) on the ^{1}H channel; the specified RF field strength is

the value at the mid-point of the ramp. Hard pulses were typically 2.5 to 2.6 µs (^{1}H), 3.4 to 3.5 µs (^{15}N) and 3.2 µs (^{13}C).

The time-shifted REDOR, ^{13}C R_1, ^{13}C $R_{1\rho}$ and ^{1}H $R_{1\rho}$ experiments used the same basic hCH correlation experiment, with the appropriate pulse sequence element as described in Figure 3.8 of reference [243]. In the ^{1}H $R_{1\rho}$ experiment, a spin-lock element was inserted after the initial ^{1}H excitation pulse. Except the ^{1}H relaxation experiment, all pulse sequences are implemented in NMRlib [274].

The REDOR experiment was performed at a MAS frequency of 44.053 kHz (rotor period of 22.7 µs). For the recoupling pulse train in the REDOR experiment, the ^{1}H-π and ^{13}C-π pulses had durations of 5 µs and 6 µs, respectively. The REDOR experiment was implemented with a shift of half of the ^{1}H pulses away from the center of the rotor period as described previously [246; 275]. The shift of the pulses was such that the shortest time interval between two consecutive ^{1}H pulses was 0.5 µs, i.e. the centers of these two consecutive ^{1}H pulses were separated by 5.5 µs. Seventeen time points were acquired, from 2 rotor periods to 36 rotor periods in steps of 2 rotor periods (one on each side of the central ^{13}C pulse). The REDOR data were collected as a series of 1D spectra.

In the ^{13}C $R_{1\rho}$ measurements, ten 1D experiments with different spin-lock durations (between 1 and 45 ms) were collected, and this was repeated for 20 different spin-lock RF field strengths, ranging from 2 kHz to 40 kHz. A ^{1}H-π pulse was applied in the center of the relaxation period to suppress cross-correlated relaxation effects [276]. In the ^{13}C R_1 measurements, ten 1D experiments with different relaxation delays (between 1 and 45 ms) were collected. In the ^{1}H $R_{1\rho}$ measurements, ten 1D experiments with different spin-lock durations (between 1 and 25 ms) were collected, and this was repeated for 25 different spin-lock RF field strengths, ranging from 1 kHz (for MPD-ub) or 2 kHz (cubic-PEG-ub) to 40 kHz.

To analyse the data, the peaks in the 1D series of spectra were integrated using in-house written python scripts. The relaxation decay profiles were fitted using a simple exponential fit. To interpret the REDOR data, a series of numerical simulations of the REDOR recoupling element was conducted using the GAMMA simulation package [277]. The simulations were performed with different values of the tensor anisotropy and biaxiality, resulting in a 2D grid of simulated time traces. The experimental data were fitted by first calculating

$\Delta S/S_{\mathrm{ref}} = (S_{\mathrm{ref}} - S_{\mathrm{rec}})/S_{\mathrm{ref}}$ for both the experiment and the simulation, where S_{rec} and S_{ref} are the signal intensities in the recoupling experiment and reference experiment (with and without ^{1}H pulses, respectively). Then, the experimental curve was compared to each simulation, by calculating a chi-square value as the sum of the squared deviations between the experimental and simulated $\Delta S/S_{\mathrm{ref}}$ divided by the squared experimental error estimate. The best-fit values of the two fitted parameters, tensor anisotropy and biaxiality, were taken to be those for which the calculated chi-square value was minimal over the simulated 2D grid. In an alternative fitting approach, we have performed the simulations with an explicit jump model, assuming that ring flips cause a change of 120° in orientation of ^{1}H-^{13}C vector. A 1D grid of these simulations was compiled, where the dipolar tensor anisotropy was varied.

Confidence intervals were obtained from Monte-Carlo simulations: the best-fit curve (relaxation decay or REDOR curve) along with experimental uncertainties (of the intensities or $\Delta S/S_{\mathrm{ref}}$ values) were used to generate a "noisy"data set, by choosing the points randomly within a normal distribution around the best-fit data point. One thousand such noisy data sets were fitted using the same procedure as described above, and the standard deviation of the fitted parameters is reported here.

The calculation of relaxation-rate constants in Figure 3.4 (D-G) used Redfield-theory-based analytical expressions, as outlined in the section 3.5.1. The $R_{1\rho}$ equations have been derived in reference [278]. The order parameter used for these calculations was set to $S = 0.661$ (unless stated otherwise), which corresponds to a two-site jump model, $S^2 = (3cos^2\varphi + 1)/4 = 0.437$, where it is assumed that the jumps occur between the two equiprobable states and cause reorientation of dipolar vector by 120° [279]. Note that $S = 0.661 = \sqrt{0}.437$ is not exactly the same as $S = 0.625$, which is the scaling of the dipolar-coupling anisotropy from the two-site jumps, used in the REDOR analysis. The latter, however, refers to a different observable and stems from the theoretical description, which also involves a tensor biaxiality.) For all calculations we have used a value of the dipolar coupling that corresponds to 1.09 Å bond length (23327 Hz tensor anisotropy [231]. The CSA tensor was assumed to be uniaxial (axially symmetric), with a value of $\Delta\delta$ =-159 ppm [280]. Note that although the validity of the Redfield theory for slow motions has been debated [237; 281], it appears to be valid for the range of rate constants

considered here [282]. More specifically, some discrepancies have been detected between Redfield-theory calculations and numerical simulations; however, where the theory appeared to produce incorrect rate constants, it turned out to be due to the fact that decays are multi-exponential. In essence, the deviations were due to misinterpretation of multi-exponential behavior. Caution is required in extracting relaxation rate constants, as discussed in refs. [283; 284].

3.3.4 MD modelling

Four MD trajectories have been analyzed to glean information on phenylalanine dynamics: those of MPD-ub (2 µs), cubic-PEG-ub (2 µs) and rod-PEG-ub (2 µs) crystals, as well as ubiquitin in solution (8 µs). For example, the MPD-ub trajectory is based on crystallographic structure 3ONS [285] and models the crystalline supercell that is comprised of 4 unit cells. In total, the simulation box contains 24 ubiquitin molecules and 8,772 SPC/E [286] water molecules. The dimensions of the box were rescaled by a factor 1.016 to reflect the expansion of the crystal at room temperature. For cubic-PEG-ub, the simulated box contains one unit cell, comprising 48 ubiquitin molecules, and for rod-PEG-ub two unit cells, containing the total of 24 ubiquitin molecules in total.

As standard for protein crystal simulations [287], the periodicity of the crystalline lattice is modeled by means of periodic boundary conditions applied to the faces of the simulation box. All crystal trajectories have been recorded in Amber 16 program [288] using ff14SB force field [273]; the solution trajectory was recorded in Amber 11 [289] using ff99SB force field [290; 291]. Other details of the MD setup can be found in our previous publications [244; 249]. MD data have been processed using python library pyxmolpp2 written in-house (available from `https://github.com/bionmr"~spbu/pyxmolpp2`). In particular, this library offers facilities to extract dihedral angles and calculate SASA.

To calculate the chemical shifts, we have used the crystallographic structures 3ONS [285], 3N30 [292] and 3EHV [293] and built the respective crystal lattices. From these lattices we carved out the fragments representing a Ub chain of interest together with the proximal chains. These fragments were subsequently fed into the chemical shift predictor SHIFTX2 [263] to predict

the phenylalanine side-chain $^{1}H^{\varepsilon}$,$^{13}C^{\varepsilon}$ chemical shifts. The obtained shifts were averaged over the ε1, ε2 pairs of atoms. In addition, we have also used the high-accuracy NMR structure 1D3Z to similarly predict the chemical shifts in solution.

To calculate the number of flips in the MD simulations, we have employed the following scheme. We defined the flip as the transition between the two states, χ_2=[60÷180°] and χ_2=[240÷360°] (cf. Figure 3.5). The very rare appearances of the rings in between of these two corridors have been ignored. In this manner we have counted all Phe4 flips observed in all of the crystal and solution simulations and similarly counted all Phe45 flips. These calculations indicate that the flip rate of Phe45 is, on average, 47 times slower than the flip rate of Phe4 in our MD simulations.

3.4 Discussion

We have shown that a selective isotope labeling scheme with proton-detected MAS NMR provide insights into phenylalanine dynamics in ubiquitin crystals. The resolution in the ^{1}H dimension allowed us to detect two distinct environments of Phe4 in the two chains in the asymmetric unit cell. Specifically, stacking of Phe4 rings between two neighboring molecules leads to an upfield proton shift, thus giving rise to a distinct signal for one of the two chains in cubic-PEG-ub. In rod-PEG-ub, which is composed of three non-equivalent chains, only one peak is seen because Phe4 points into the solvent in all chains. Phe45, unobserved in all three crystals, is presumably broadened beyond detection by the slow flips, as suggested by MD simulations. Using the experimentally determined time scale of ca. 10-20 ns for the flips of Phe4 (Figure 3.4 (E)) and the MD-derived factor by which Phe45 is slower than Phe4, ca. 50, we estimate that Phe45 ring flips occur on a time scale of about 1 µs.

We have calculated proton transverse relaxation rate $R_{1\rho}$ (using a 4-spin system, see section 3.5 for details), and carbon $R_{1\rho}$ as a function of the time scale of dynamics, to estimate how rapidly the spin magnetization would decay if ring flips occur on such long time scales. These data (Figure 3.4 (E, G)) show that the coherence life times under a spin lock are a fraction of millisecond.

Such a rapid decay means that the signal is expected to be very broad, and during the coherence transfer steps in an hCH experiment it should die off before detection. In a simple proton single-pulse excitation 1D spectrum, the aromatic signals overlap with the amide signals, making it impossible to detect Phe45. Deuterating fully all amide sites, i.e. using 100 % deuterated buffer, is not possible as even the precipitation agent, MPD, brings in about 20 % ^{1}H into the solvent.

It is noteworthy that Phe45 has been detected in one MAS NMR study of MPD-ub, although at lower temperature (273 K) and using carbon detected experiments [262]. While we can only speculate about the reasons, a possibility is that Phe45 in MPD-ub has ring flips that are much slower than in other crystals, e.g. in hundreds of µs. This conjecture finds some support in our MD simulations of MPD-ub crystal, where Phe45 shows only one single flip (see Figure 3.5 (B)). If so, then lowering the temperature to 273 K may further slow down Phe45 flips, bringing them to millisecond territory [294]. That would create the conditions for Phe45 signal to become observable again, cf. Figure 3.4 (E, G). In this connection it should be mentioned that under the conditions of fast MAS used in our study, the lowest temperature that we could achieve was ca. 15-20 °C, and we did not observe the Phe45 peak in these trial experiments (not shown).

Of note, the B factors of the δ and ε carbons that sense flips are comparable for Phe4 and Phe45 in each of the crystals (see Figure 3.10). It is of course not surprising that the strong differences that we find in NMR data are not seen by X-ray crystallography: the difference between Phe4 and Phe45 is the time scale of flips, and crystallography cannot see flips nor their time scale; at cryogenic temperatures they are furthermore expected to be frozen out.

We sought to identify the origin of the large difference in ring-flip rates between Phe4 and Phe45. Given that Phe4 and Phe45 are not buried in the hydrophobic core, but are positioned on the outside of the ubiquitin molecule, a parameter that likely has an influence is the solvent accessible surface area (SASA). We reasoned that SASA should reflect the void (i.e. in this case water-filled) volume available to the phenyl ring. Another factor that might influence the flip rate is the arrangement of phenylalanine side chain relative to the backbone, i.e. its rotameric state.

Figure 3.6 summarizes structural information and MD data on rotameric states and SASA of the two phenylalanine residues in ubiquitin. The heat

maps in this figure show that Phe4 and Phe45 populate two distinct rotameric states, possibly offering an explanation for the observed differences in ring-flip rate constants. To test this proposition, we looked at surface-exposed Phe residues in MD trajectories of several unrelated proteins. In doing so, we found phenylalanines that belong to the same rotameric state as Phe4 but do not experience any flips, as well as other phenylalanines that belong to the same rotameric state as Phe45 but undergo frequent flips (Figure 3.11). This observation rules out the possibility that the rotameric states can explain the differences in ring flips. Figure 3.6 also shows SASA distributions as obtained from our MD simulations, as well as SASA values from solution and crystal structures (in addition, SASA variations on per-molecule basis are illustrated in Figure 3.12). No clear correlation is found between the SASA characteristics and the observed ring-flip rates. Thus, we are led to conclude that the flip rates are likely controlled by a mix of structure and dynamics factors, involving phenylalanine residues and their immediate surroundings, that remain to be fully identified.

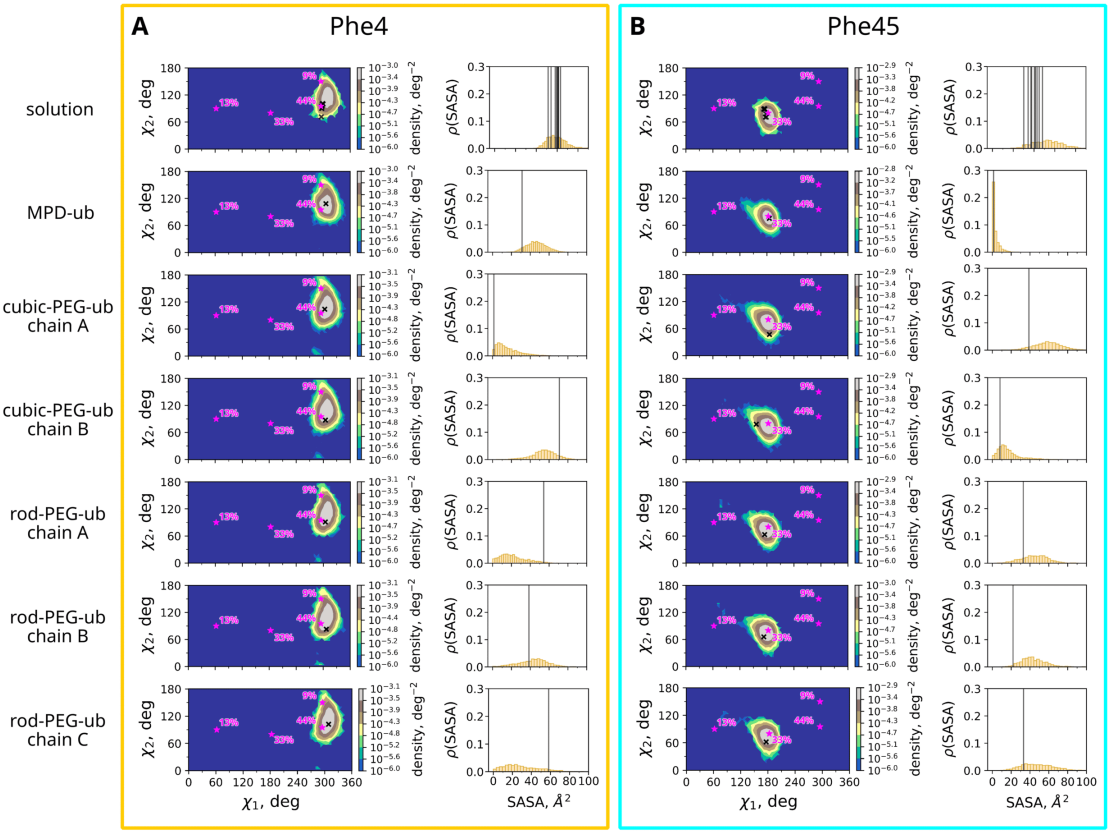


Figure 3.6 — Rotameric states and solvent-accessible surface areas for residues (A) Phe 4 and (B) Phe45 from MD simulations and various structural data. Respective crystal and solution species are identified on the left side of the plot. First column: heat maps showing (χ_1,χ2) probability density distribution

for Phe4 side chain as obtained from our MD data. Shown with black crosses are the values of χ_1, χ_2 according to solution NMR (1D3Z, 10 conformers) or crystallographic (3ONS, 3N30, 3EHV) structures. In addition, canonical phenylalanine rotamers according to ref. [295] are indicated with magenta stars along with their respective frequencies of occurrence. The χ_1 and χ_2 range is [0÷360°], same as in Figure 3.5. Second column: histograms showing SASA distributions for Phe4 residue as obtained from our MD data. Also shown are Phe4 SASA values from the solution NMR and crystallographic structures (black vertical lines). Third and fourth columns: the same data for ubiquitin residue Phe45.

Interestingly, while a previous study has found overall rocking motion in cubic-PEG-ub crystal, which leads to strong NERRD effects and overall higher $R_{1\rho}$ relaxation rates, we do not find significantly different relaxation for the Phe carbons of that particular crystal, and any NERRD effect seen therein is weak, if not absent (Figure 3.4 (C)). It is possible that the effects of rocking motion are masked in this case by ring flips.

Overall, our study reveals that for ubiquitin the crystal packing has little effect on ring-flip dynamics: for the observed Phe4, the flip rate constants are nearly the same, and in all crystals Phe45 appears to be broadened beyond detection, which we ascribe to much slower flips. It will be interesting to probe the effects of crystalline packing on ring flips of aromatic residues buried in the hydrophobic protein core, where «breathing motions» are thought to be required for ring flips. We expect that the said breathing motions are largely insensitive to the soft restraining effect of crystal contacts. The methodology we have presented here, which combines selective labeling and a suite of MAS NMR experiments, is well suited to address this question.

3.5 Supporting Information

3.5.1 Calculation of $R_{1\rho}$ relaxation

The ^{1}H $R_{1\rho}$ relaxation can be described by a sum of heteronuclear dipolar ^{1}H-^{13}C ($R_{1\rho}{}^{HC}$), homonuclear dipolar ^{1}H-^{1}H ($R_{1\rho}{}^{HH}$) and the ^{1}H CSA ($R_{1\rho}{}^{CSA}$) contributions [296]:

$$\begin{aligned} R_{1\rho}^{HC} = \frac{1}{4} {d_{HC}}^2 \Big\{ \frac{1}{3} \sin^2(\beta_{eff}) \Big[& J(\omega_e - 2\omega_r) + 2J(\omega_e - \omega_r) + 2J(\omega_e + \omega_r) + \\ & + J(\omega_e + 2\omega_r) + 9J(\omega_C) \Big] + \\ & + \frac{1}{4}(3 + \cos(2\beta_{eff})) \Big[3J(\omega_H) + 6J(\omega_H + \omega_C) + J(\omega_H - \omega_C) \Big] \Big\} \end{aligned} \tag{3.1}$$

$$\begin{aligned} R_{1\rho}^{HH} = \frac{3}{4} {d_{HH}}^2 \Big\{ & \frac{1}{8} \sin^2(2\beta_{eff}) \Big[J(\omega_r - \omega_e) + J(\omega_r + \omega_e) \Big] + \\ & + \frac{1}{2} \sin^4(\beta_{eff}) \Big[J(\omega_r - 2\omega_e) + J(\omega_r + 2\omega_e) \Big] + \\ & + \frac{1}{16} \sin^2(2\beta_{eff}) \Big[J(2\omega_r - \omega_e) + J(2\omega_r + \omega_e) \Big] + \\ & + \frac{1}{4} \sin^4(\beta_{eff}) \Big[J(2\omega_r - 2\omega_e) + J(2\omega_r + 2\omega_e) \Big] + \\ & + \frac{1}{4}(7 - 3\cos(2\beta_{eff})) \Big[J(\omega_H) \Big] + \\ & + \frac{1}{2}(5 + 3\cos(2\beta_{eff})) \Big[J(2\omega_H) \Big] \Big\} \end{aligned} \tag{3.2}$$

$$\begin{aligned} R_{1\rho}^{CSA} = \frac{1}{3} \omega_H^2 \Delta\sigma^2 \Big\{ & \frac{1}{9} \sin^2(\beta_{eff}) \Big[J(\omega_e - 2\omega_r) + 2J(\omega_e - \omega_r) + \\ & + 2J(\omega_e + \omega_r) + J(\omega_e + 2\omega_r) \Big] + \\ & + \frac{1}{4} (3 + \cos(2\beta_{eff}) \Big[J(\omega_H) \Big] \Big\} \end{aligned} \tag{3.3}$$

where, d_{HC} and d_{HH} are the hetero- and homonuclear dipolar couplings given by $d_{12} = -\frac{\mu_0}{4\pi} \frac{\gamma_1 \gamma_2 \hbar}{r_{12}^3}$ with the vacuum permeability μ_0, the gyromagnetic ratios γ_i, Planck's constant $\hbar$ and the internuclear distance r_{12}. $\Delta\delta = \frac{3}{2}(\delta_{zz} - \delta_{iso})$

is the chemical shift anisotropy, ω_e is the amplitude of the effective field, β_{eff} is the tilt angle of the effective field, ω_r is the MAS frequency, and ω_H and ω_C are the ^{1}H and ^{13}C Larmor frequencies, respectively. The spectral densities $J(\omega)$ are defined as $J(\omega) = \frac{2}{5}(1 - S^2)\frac{\tau_c}{1 + (\omega\tau_c)^2}$, with the order parameter S and the correlation time τ_c.

For the homonuclear dipolar relaxation of an H$^\varepsilon$ proton (Figure 3.4 (G)), we considered a four-spin system inspired by the environment of Phe45 in the current deuterated, selectively ^{1}H,^{13}C-labeled sample. This system contains the H$^\varepsilon$ proton of interest, its directly bound C$^\varepsilon$, another H$^\varepsilon$ proton and the closest H^N amide proton in the backbone (Thr66). The distances between the nuclei are 1.09 Å for C$^\varepsilon$-H$^\varepsilon$, 4.3 Å for H$^\varepsilon$-H$^\varepsilon$ and 3.0 Å for C$^\varepsilon$-H^N. In addition to the dipolar couplings that result from these spin pairs, we have taken into account the H$^\varepsilon$ CSA. The proton chemical shift anisotropy $\Delta\delta$ was set to 9 ppm [297]. The relaxation rates were calculated assuming a magnetic field corresponding to 600 MHz proton Larmor frequency, 44.053 kHz MAS, an effective field of 80 kHz and an effective tilt angle of 90°. The order parameters S were set to 0.661 for the heteronuclear ^{13}C$^\varepsilon$-^{1}H$^\varepsilon$ relaxation, 1.0 for the homonuclear ^{1}H$^\varepsilon$-^{1}H$^\varepsilon$ relaxation and 0.5 for the homonuclear ^{1}H$^\varepsilon$-^{1}H^N relaxation [279].

3.5.2 Including the effect of motional modes other than ring flips

In addition to ring flips, ^{13}C$^\varepsilon$ spin relaxation is also sensitive to other motional modes, such as small-amplitude axial fluctuations involving χ_1 and χ_2. To estimate the effect of these other modes on the outcome of our analyses, Figure 3.4 (E), we have applied the following scheme. In MPD-ub trajectory, we have identified 6 ubiquitin molecules where Phe4 ring does not experience any flips. For these six sites, we have calculated the average ^{1}H$^\varepsilon$-^{13}C$^\varepsilon$ dipolar correlation function $g(\tau)$, shown as black curve in Figure 3.8. The curve has a familiar shape, decaying toward the plateau value of 0.84; similar shapes are found for backbone sites, where dynamics is also limited to small-amplitude fluctuations [298]. As is generally the case, the decay of $g(\tau)$ has a multi-exponential character [299]. Accordingly, the function can be fitted with a combination of several exponentials in the spirit of the extended Lipari-Szabo

model [300]. Here we have found that four exponential terms are necessary and sufficient to fit $g(\tau)$. For the purpose of our analyses, we have assumed that there are four independent motional modes characterized by their respective order parameters S_i^2 and correlation times τ_i. We have further assumed that the net correlation function is a product of the four mode-specific functions [301]:

$$\begin{aligned} g_{fit}(\tau) = ((1-S_1^2)exp(-\tau/\tau_1)+S_1^2) \times ((1-S_2^2)exp(-\tau/\tau_2)+S_2^2)\times \\ \times((1-S_3^2)exp(-\tau/\tau_3)+S_3^2) \times ((1-S_4^2)exp(-\tau/\tau_4)+S_4^2) \end{aligned} \tag{3.4}$$

where $g(\tau)$ correlation function with the ansatz from Equation (3.4) is illustrated in Figure 3.8 (dashed green curve in the plot). The obtained time constants, $\tau_1 = 0.98$ ps, $\tau_2 = 32$ ps, $\tau_3 = 0.59$ ns, $\tau_4 = 9.2$ ns are different by an order of magnitude, thus confirming the statistical independence of the respective motional modes. The first time scale, τ_1, is extremely short and, in fact, determined by the time interval that is used to record protein coordinates during the MD simulation. This time scale is clearly not relaxation-active and, therefore, the expression for $g_{fit}(\tau)$ can be simplified to:

$$\begin{aligned} g_{red}(\tau) = S_1^2 \times ((1-S_2^2)exp(-\tau/\tau_2)+S_2^2) \times ((1-S_3^2)exp(-\tau/\tau_3)+S_3^2)\times \\ \times ((1-S_4^2)exp(-\tau/\tau_4)+S_4^2) \end{aligned} \tag{3.5}$$

Furthermore, the term S_1^2, which pertains to sub-picosecond vibrations and librations, must be already factored into the interaction constants (dipolar and CSA) used in the data analyses. Therefore, the relevant portion of the correlation function is reduced to:

$$\begin{aligned} g_{red}(\tau) = ((1-S_2^2)exp(-\tau/\tau_2)+S_2^2) \times ((1-S_3^2)exp(-\tau/\tau_3)+S_3^2)\times \\ \times ((1-S_4^2)exp(-\tau/\tau_4)+S_4^2) \end{aligned} \tag{3.6}$$

Using this MD-derived correlation function, we now construct a more general correlation function, which additionally accounts for the effect of ring flips:

$$\begin{aligned} G(\tau) = ((1-S_2^2)exp(-\tau/\tau_2)+S_2^2) \times ((1-S_3^2)exp(-\tau/\tau_3)+S_3^2)\times \\ \times ((1-S_4^2)exp(-\tau/\tau_4)+S_4^2) \times ((1-S_{flip}^2)exp(-\tau/\tau_{flip})+S_{flip}^2) \end{aligned} \tag{3.7}$$

In this expression, the flip order parameter is $S^2_{flip} = 0.437$ [279], the values of S^2_2, τ_2, S^2_3, τ_3 and S^2_4, τ_4 are fixed according to the MD fitting results (Figure 3.8), and τ_{flip} is the sole variable. One can notice that in our case $\tau_2 \ll \tau_3 \ll \tau_4, \tau_{flip}$, which permits the following simplification:

$$J(\omega) = \frac{2}{5}[(1 - S^2_2)\frac{\tau_2}{1 + (\omega\tau_2)^2} + S^2_2(1 - S^2_3)\frac{\tau_3}{1 + (\omega\tau_3)^2} + S^2_2 S^2_3 \xi] \tag{3.8}$$

$$\begin{aligned} \xi = (1 - S^2_4)(1 - S^2_{flip})\frac{\tau'}{1 + (\omega\tau')^2} + (1 - S^2_4)S^2_{flip}\frac{\tau_4}{1 + (\omega\tau_4)^2} + \\ + S^2_4(1 - S^2_{flip})\frac{\tau_{flip}}{1 + (\omega\tau_{flip})^2} \end{aligned} \tag{3.9}$$

$$\tau' = (\frac{1}{\tau_4} + \frac{1}{\tau_{flip}})^{-1} \tag{3.10}$$

Note that this simplification is valid only for our specific case, where the time scales τ_2, τ_3 and (τ_4, τ_{flip}) are separated by at least an order of magnitude, see Figure 3.8 and Table 8.

Equations (3.8) - (3.10) have been used to re-interpret our experimental results, arriving at Figure 3.9. Comparing this latter to Figure 3.4 (E) in the main text, one can appreciate the significance of motional modes other than ring flips and quantify the bias in determination of τ_{flip}.

Finally, note that MD-derived parameters of local dynamics are not always accurate. Furthermore, different motional modes underlying the correlation function may not be completely independent, but rather may prove to be partially correlated. In addition, the relevant constants, e.g. $r_{CH} = 1.09$ Å, are also not very accurately known and are a subject of debate focusing on the effect of vibrational averaging [302; 303]. Thus, the results in Figure 3.9 are, at best, a semi-quantitative estimate.

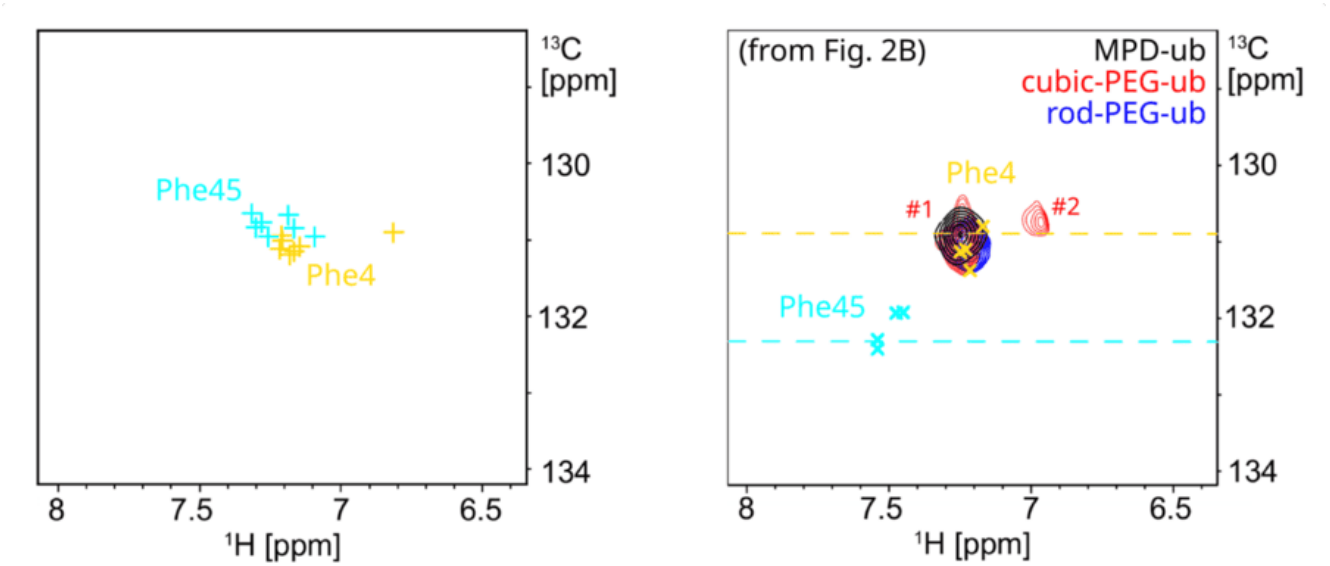


Figure 3.7 — Chemical-shift predictions for $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ spin pairs in Phe4 (colored yellow) and Phe45 (colored cyan) using the program SHIFTX2. It appears that SHIFTX2 correctly predicts the similarity of Phe45 chemical shifts in different crystal forms and in solution. Furthermore, it correctly predicts the Phe4 shifts, including the resonance from chain A in the cubic-PEG-ub crystal, which is shifted due to the intermolecular stacking of Phe4 rings (see Fig 3.2 (C)). However, SHIFTX2 apparently fails to accurately predict the absolute value of Phe45 shifts as can be deduced from the experimental evidence, see Fig 3.2 (B).

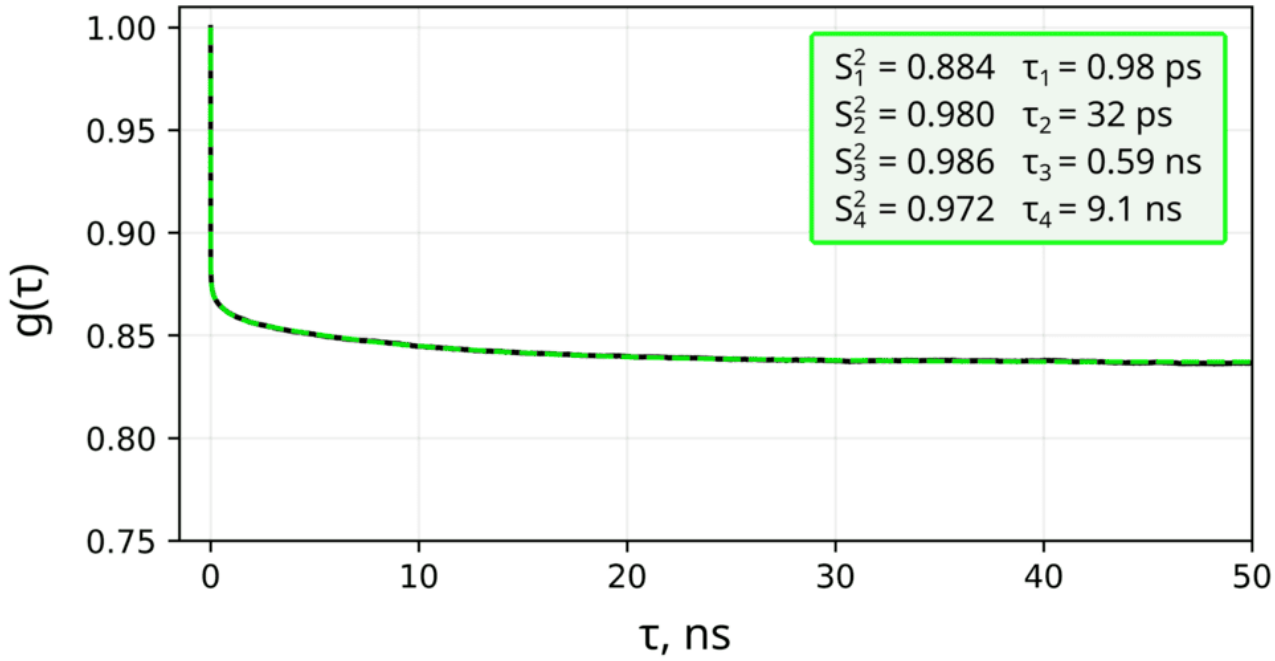


Figure 3.8 — $^{1}H^{\varepsilon}$-$^{13}C^{\varepsilon}$ dipolar correlation function reflecting phenylalanine motions other than ring flips. The data are from six Phe4 residues that do not experience any flips in our MPD-ub simulation (see Figure 3.5 (A)). The time correlation function $P_2(\cos\theta)$ (black curve) has been fitted using the 4-exponential decay function parameterized in the spirit of model-free model, Equation (3.4) (dashed green curve). The best-fit order parameters and correlation times are listed in the inset. The effect of these non-flip modes on our analyses of spin relaxation rates is illustrated in Figure 3.9.

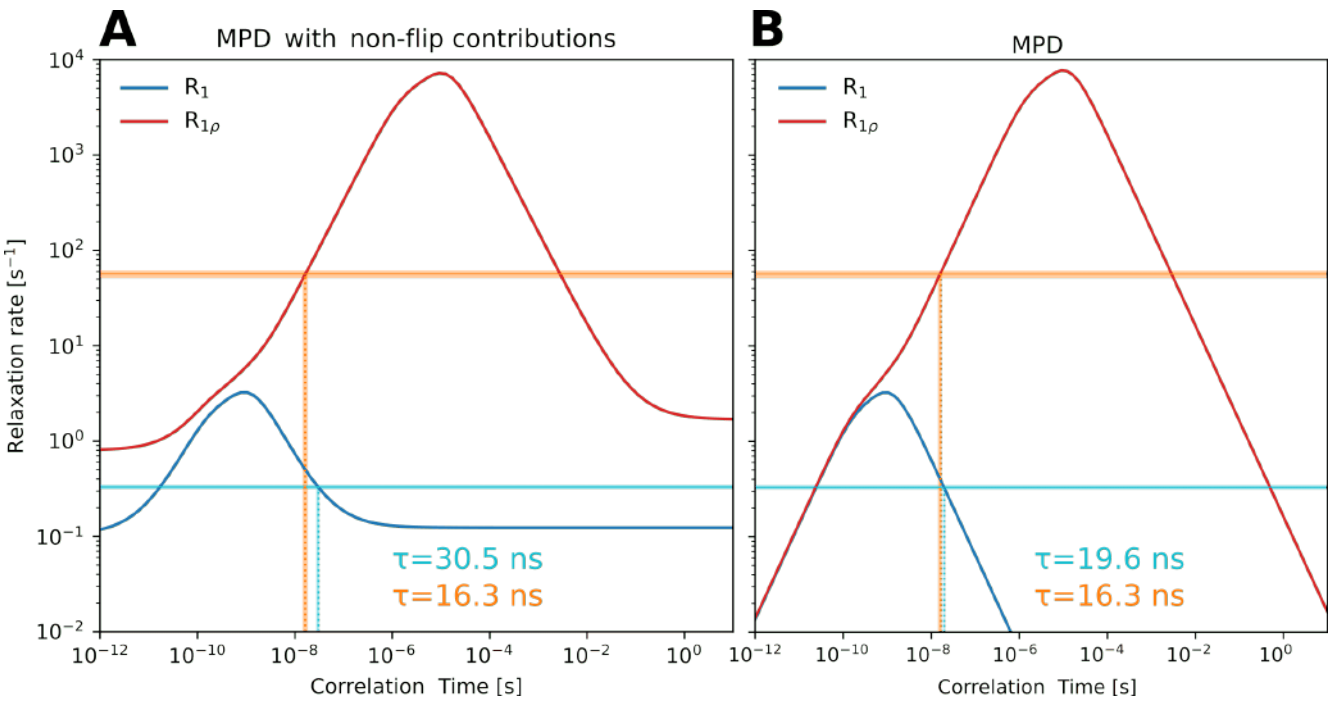


Figure 3.9 — Investigation of the contributions of non-flip motions to the ^{13}C relaxation rate constants and their effect on the extracted ring-flip correlation times. (A) Re-analysis of ^{13}C $R_{1\rho}$ and ^{13}C R_1 data from the MPD-ub crystals, which additionally accounts for non-flip motions as listed in Figure 3.8. We used the experimentally observed relaxation rate constants to re-determine the ring-flip time constant τ_{flip}. The two profiles in the plot are calculated using Equations ((3.8), (3.10)), see section 3.5.2. (B) The original analysis of ^{13}C $R_{1\rho}$ and ^{13}C R_1 data from the MPD-ub crystals, limited to ring-flip dynamics (reproduced from Figure 3.4 (E)). Comparing the results in panels (A) and (B) suggests that taking into consideration non-flip dynamics causes only relatively small change in the determined ring-flip rates.

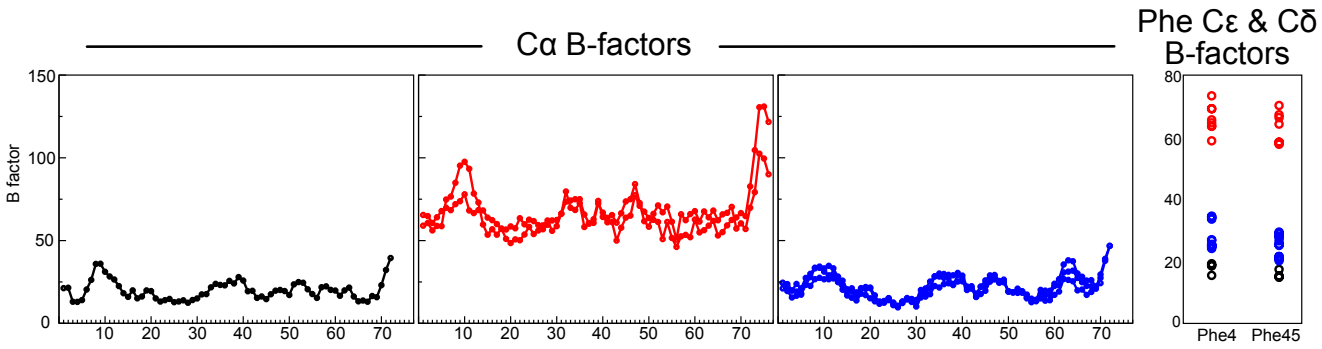


Figure 3.10 — B factors from the crystal structures of MPD-ub (black, PDB 3ONS), cubic-PEG-ub (red, PDB 3N30) and rod-PEG-ub (blue, PDB 3EHV). The left three panels show the C^{α} B factors, and the right panel shows those of the two C^{δ} and C^{ε} sites of the Phe rings. Note the large offset of the B factors of cubic-PEG-ub, which we ascribe to overall rocking motion of the protein in the crystal as investigated in ref. [249].

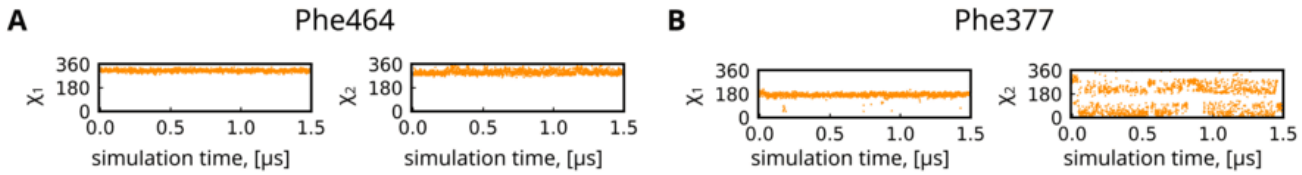


Figure 3.11 — Time traces of side-chain torsional angles χ_1 and χ_2 for surface-exposed residues (A) Phe464 and (B) Phe377 from MD simulation of receptor-binding domain from SARS-CoV-2 spike protein in complex with the mini-protein LCB1 (PDB ID: 7JZU [304]). The 1.5-μs trajectory of this complex was recorded in Amber ff14SB force field using TIP4P-Ew water. Residue Phe464 is in the same rotameric state as Phe4 in ubiquitin, m-85° according to Lovell et al. [295]; it is also highly solvent-exposed as evidenced by SASA of 72 Å^2 (calculated from the MD trajectory). However, unlike Phe4 in ubiquitin, residue Phe464 shows no flips in our simulations, $\tau_{div} \geqslant 1500$ ns. At the same time, residue Phe377 is in the same rotameric state as Phe45 in ubiquitin, t80°, and solvated at about the same level as Phe45, with SASA of 57 Å^2. Yet, unlike Phe45 in ubiquitin, residue Phe377 engages in frequent flips, $\tau_{div} = 10$ ns. These two examples demonstrate that the rotameric state of phenylalanine side chain is not a unique determinant of the flip rate.

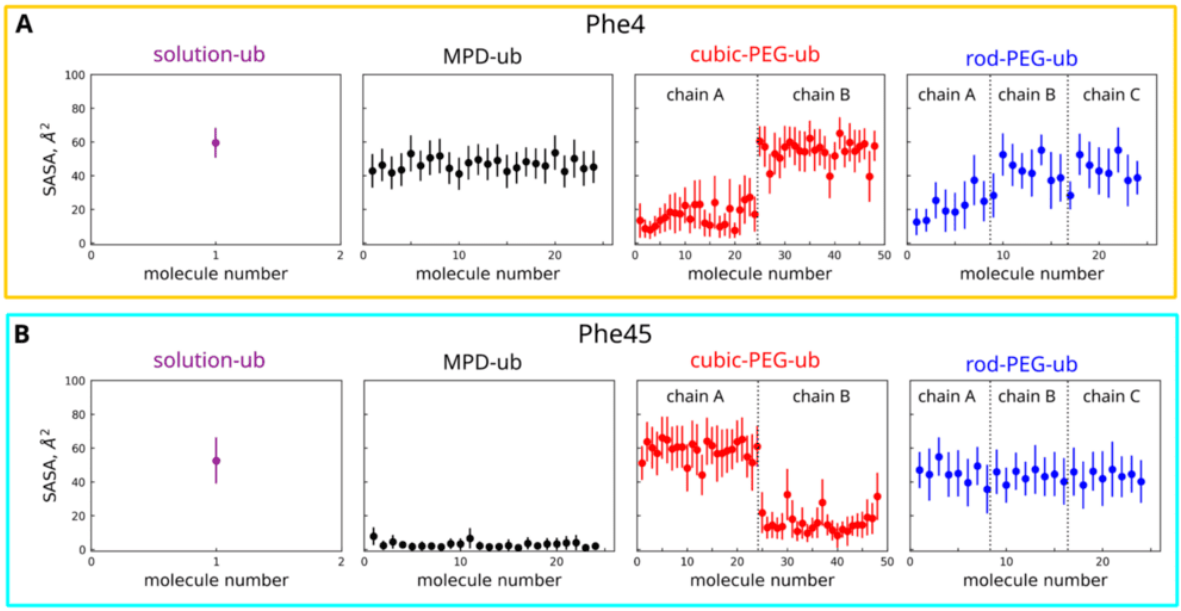


Figure 3.12 — The solvent-accessible surface areas (SASA) of the two phenylalanine residues averaged over 2-μs MD simulations of the three investigated ubiquitin crystals and 8-μs MD simulation of ubiquitin in solution. Of note, the SASA values for Phe4 and Phe45 do not show any significant correlation with the flip rates for these two residues (see Table 8). The color coding is the same as elsewhere: black (MPD-ub), red (cubic-PEG-ub), blue (rod-PEG-ub) and magenta (solution form). The MPD-ub simulation box contains 24 ubiquitin molecules that are all classified as chain A. The cubic-PEG-ub simulation box contains 48 ubiquitin molecules equally divided between chain A (molecules 1-24) and chain B (molecules 25-48). The rod-PEG-ub simulation box contains 24 ubiqutin molecules equally divided between chain A (molecules 1-8), chain B (molecules 9-16) and chain C (molecules 17-24).

Table 8 — Measured $^{13}C^{\varepsilon}$ R_1 and $R_{1\rho}$ rate constants for Phe4 in different crystal forms of ubiquitin and the corresponding calculated correlation times τ. The $R_{1\rho}$ rate constants are from the measurements at 30 kHz spin lock.

Crystal	R_1 [s^{-1}]	τ_{R_1} [ns]	$R_{1\rho}$ [s^{-1}]	$\tau_{R_{1\rho}}$ [ns]
MPD-ub	0.33 ± 0.01	$18.7\ ^{+1.4}_{-0.2}$	56.5 ± 4.1	$16.3\ ^{+1.2}_{-2.2}$
cubic-PEG-ub, peak (#1)	0.47 ± 0.05	$13.2\ ^{+2.0}_{-0.9}$	53.5 ± 5.1	$15.2\ ^{+1.1}_{-2.0}$
cubic-PEG-ub, peak (#2)	0.61 ± 0.05	$10.7\ ^{+0.8}_{-0.8}$	65.8 ± 6.2	$18.7\ ^{+1.4}_{-2.5}$
rod-PEG-ub	0.55 ± 0.03	$11.5\ ^{+0.9}_{-0.8}$	32.5 ± 0.7	$10.0\ ^{+0.1}_{-0.2}$

Table 9 — The characteristic times of Phe flips in ubiqutin according to MD simulations. The time τ_{div} is the length of the trajectory divided by the number of flips. The time τ_{corr} is the flip correlation time, which has been calculated along the lines of Redfield theory treatment. Specifically, we have extracted the time points of all Phe flips from a given MD trajectory and constructed a pseudo-trajectory where dipolar vector (e.g. H^{ε}-C^{ε}) changes its orientation by 120° with each flip. The obtained pseudo-trajectory reproduces Phe ring flips, while eliminating other motional modes such as fast axial fluctuations. The pseudo-trajectory was used to compute the correlation function $P_2(\cos(\theta))$, which was subsequently fitted with $g(\tau) = S^2 + (1 - S^2)exp(-\tau/\tau_{corr})$, where the order parameter is $S^2 = 0.437$, according to the relevant 2-site jump model [279].

	Phe4		Phe45*	
	τ_{div} (ns)	τ_{corr} (ns)	τ_{div} (ns)	τ_{corr} (ns)
solution	98	66	4000	-
MPD-ub	906	762	48000	-
cubic-PEG-ub chain A	215	170	4000	-
cubic-PEG-ub chain B	118	174	>48000	-
rod-PEG-ub chain A	236	217	4000	-
rod-PEG-ub chain B	140	242	>16000	-
rod-PEG-ub chain C	334	245	8000	-

Conclusion

In the last thirty years life scientists came to recognize that disordered protein sequences play a major role in cell activity in both normal and pathological states. The function of disordered proteins is typically effected through their interactions with partner proteins. The main scenario of such interactions involves the so-called "folding upon binding" mechanism. This mechanism assumes that disordered proteins acquire structure when binding to their structured targets; in this case, the target serves as a template to facilitate the folding of the initially disordered chain and to stabilize the formed structure. However, the repertoire of different binding modes involving disordered proteins is not limited to folding upon binding. For example, the protein ligand may partially retain its disordered character upon binding to its target, thereby forming a «fuzzy complex». Furthermore, a number of cases have been reported where two disordered proteins remain structureless after forming a tight complex. In our recent work, we have proposed a general classification for different binding modes involving disordered proteins [305].

Playing devil's advocate, one may ask a number of critical questions: what is the significance of a disordered protein *per se* so long as its function is accomplished through ordering upon binding? Should not we assume that disordered protein is merely a random polymer coil and focus instead on characterization of the fully structured complex? To answer these questions, recall that disordered proteins possess certain structural propensities, which are most clearly manifested in a form of residual structure (e.g. some regions in the disordered chain often have a tendency to form short-lived α-helices). Usually it is such elements of residual structure that enable the binding of the disordered protein to its structured target; in this case, it can be said that the binding occurs according to the "lock and key" model as opposed to the "induced fit" model [306]. Consequently, the presence of the elements of residual structure in the disordered protein has a direct effect on the binding parameters (k_{on}, K_d). In turn, even small changes in k_{on} and K_d can have a significant influence on cell activities provided that the key signaling cascades are affected. Hence, it is the residual structure that deserves thorough investigation; in other words, we are primarily interested in what distinguishes the disordered protein from a random coil.

It is evident that standard methods of structure modeling are ill-suited for disordered proteins. As an obvious alternative, molecular dynamics models (i.e. MD trajectories) can be used to represent IDPs. Yet this approach faces a number of hurdles. As it happens, MD models are usually not very accurate. It is known that during the course of long MD simulations protein molecules tend to experience the effect of structural drift; the implication is that MD force fields in most cases are unable to reproduce the global free-energy minimum corresponding to the experimentally determined protein structure [307]. For disordered proteins, the situation is exacerbated since there is not even a local minimum that could stabilize the simulated structure; this situation promotes an extensive structural drift. One should also bear in mind that MD force fields have been historically developed to model globular proteins. In practice, using such traditional force fields to model IDPs lead to some distinctive biases – e.g. an extended protein can be transformed into an exceedingly dense globule during the course of the simulation [18].

Thus, before using MD simulations to model disordered proteins one needs to address the deficiencies of the existing force fields. Solving this problem requires an extensive validation of the MD models using the relevant experimental data. This kind of strategy is actively pursued by a number of research groups, but their efforts remain rather disjointed. The tests are conducted on a host of proteins and peptides using different force fields in conjunction with multiple water models. Furthermore, the experimental datasets employed in these tests are usually very limited in scope and vary from one study to another. At the same time, the MD-based calculations of measurable parameters are often carried out using simplified or even crude formulas, which largely negates the value of comparative analyses. As a result, there is currently no consensus protocol for MD modeling of the partially or fully disordered proteins, and the conclusions from many validation studies are often at odds with each other.

In our opinion, to achieve a progress in this area the scientific community should focus on a few select model systems. One of them can be the peptide N-H4 that has been discussed in the chapter 1 of this dissertation. For each of the selected systems, an extensive set of experimental data should be collected and then used to systematically test different force fields (including water models). The results of these tests should be used to identify those force fields (water

models) that are best suited for modeling of disordered proteins and, ultimately, to develop the next-generation of force fields. Such systematic studies can greatly benefit from multiple research groups joining forces to explore a range of modeling options. It is worth mentioning that the author of this dissertation has an experience of organizing such an international research consortium, aiming to investigate the effect of posttranslational modifications on histone tail dynamics (the study is currently ongoing and is not covered in this text).

Regarding the quality of the MD models, there is room for improvement of the existing conventional force fields. Work to optimize the parameters of various force fields is being done in a number of research groups, marked with some considerable successes [308]. Yet, sooner or later this line of development will reach its limit, thus prompting the research community to turn to more elaborate and complex modeling algorithms. This transition would require a significant increase in computational power; however, the ongoing progress in computing technologies can (at least in part) satisfy these needs. As an example of such more sophisticated approach to MD modeling one can cite the use of polarizable force fields. It should be mentioned, though, that the use of polarizable force fields has not so far led to any notable breakthroughs in the modeling of biomolecular systems.

In this connection, one should bear in mind that applications of molecular dynamics to conformationally disordered systems call for ultra-long MD simulations, which are necessary to adequately sample the immense conformational diversity of such systems. The need for long trajectories is especially pressing when the goal is to use the MD data for calculation of paramagnetic relaxation rates. Indeed, as discussed in chapter 2 of this thesis, the calculations of PREs are extremely demanding in terms of convergence. In many cases, the problem of convergence can be mitigated by using the specialized MD modeling schemes. For example, the computational scheme known as metadynamics can be very useful when modeling side-chain dynamics in globular proteins, as investigated in chapter 3.

Finally, it should be noted that further progress in this area should involve more diverse and rich molecular context. For instance, this includes posttranslational modifications of histone tails, which have already been mentioned above. Along these lines one could also simulate molecular assemblies with increased number of components. For example, as a part of the nucleosome

project we have obtained some preliminary results for the MD model of crystalline tetranucleosome, offering some insight into the role of histone tails in chromatin packing. Yet another example of our ongoing research in this area is MD modeling of the interaction between the pioneer transcription factor Cfb1 and the nucleosome core particle, where the flexible tail of histone H3 serves as a point of contact for Cfb1.

Perhaps, the most striking example of in silico studies in this area is the computer-assisted search for therapeutic ligands to target disordered peptide chains. Many attempts to develop small-molecule ligands of IDPs have met with only limited success, even though some promising results have been reported [309]. It can be assumed that the very nature of disordered proteins is not conducive to small-molecule binding (in contrast to structured proteins, where small-molecule ligands are common). At the same time, it is typical of IDPs to bind structured proteins that act as their physiological targets. This opens an avenue for development of de novo proteins that are capable of engaging IDPs. The first demonstration of this new approach has been reported by the research team led by D. Baker in the summer of 2024 [310]. It is symbolic that on the day when these lines are written the Nobel Prize has been awarded to D. Baker for his fundamental contributions to the field of protein engineering. It can be anticipated that computer methods to design protein ligands of disordered protein targets should become one of the powerful instruments in healthcare technology of the future. The development of accurate IDP models and their validation using suitable experimental data, as described in this dissertation study, are the integral part of the efforts in this direction.

Acronyms

BBI	broadband inverse probe
BRET	bioluminescence resonance energy transfer
CSA	chemical shift anisotropy
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ESR	electron spin resonance
FRET	Förster energy resonance transfer
IDPs	Intrinsically disordered proteins
IDPR	Intrinsically disordered regions
HSQC	Heteronuclear Single Quantum Coherence
HMR	hydrogen mass repartition scheme
MAS	Magic Angle Spinning
MD	Molecular dynamics
MSD	mean square deviation
NCP	nucleosome core particle
N-H4	peptide SGRGKGGKGLGKGGAKRHRKVLRDN
NOE	nuclear Overhauser effect
NPT	thermodynamic ensemble with the constant number of particles, pressure and temperature
NVE	thermodynamic ensemble with the constant number of particles, volume and energy
OPC	optimal point charges
PDB	Protein Data Bank
PFG	Pulsed Field Gradient
PRE	Paramagnetic Relaxation Enhancement
RMSD	root mean square deviation
SASA	solvent-accessible surface area
SAXS	small-angle X-ray scattering
TIP3P	transferable intermolecular potential, 3 point
TIP4P-Ew	transferable intermolecular potential, 4 point, Ewald
TIP4P-D	transferable intermolecular potential, 4 point, dispersion

Acknowledgements

I would like to express my deep appreciation to my scientific advisor, Nikolai Skrynnikov, and my colleagues in the Laboratory of Biomolecular NMR, Ivan Podkorytov and Sergei Izmailov, for many years of support and the opportunity to work in a world-class research group. I am particularly indebted to Ivan Podkorytov for his advice and encouragement in working on this dissertation. My research has been greatly facilitated by the atmosphere of the open scientific discussion and camaraderie in the Laboratory of Biomolecular NMR owing to its former and current members: Elena Bushmanova, Polina Vaganova, Aleksandr Kovalenko, Svetlana Korban, Aleksandra Levkina, Dmitry Luzik, Oleg Mikhailovskii, Olga Rogacheva, Vladislav Salikov, Irina Tyuryaeva and Boris Kharkov. I am also deeply thankful to the leaders of the research groups, whose collaboration made a big contribution to the success of this work: Paul Schanda, Chrstopher Jaroniec, David Case, Alexey Onufriev and Alexey Shaitan.

I wish to acknowledge the contribution from a number of Resource Centers in the Research Park of SPbU that provided resources for many computations and experiments described in this dissertation. These include the «Computing Centre», «Magnetic Resonance Research Centre», «Chemical Analysis and Materials Research Centre », «Centre for Molecular and Cell Technologies», «Centre for X-ray Diffraction Studies», «Thermogravimetric and Calorimetric Research Centre», as well as the Cryogen section within in the SPbU Research Park. My special thanks are extended to the Faculty of Physics at SPbU and the Department of Nuclear Physics Research Methods.

Bibliography

1. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis / J. C. Kendrew [et al.] // Nature. — 1958. — Mar. — Vol. 181, no. 4610. — P. 662—666.

2. Wüthrich, K. NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor / K. Wüthrich, G. Wagner // FEBS Letters. — 1975. — Feb. — Vol. 50, no. 2. — P. 265—268.

3. Bennett, W. S. Glucose-induced conformational change in yeast hexokinase. / W. S. Bennett, T. A. Steitz // Proceedings of the National Academy of Sciences. — 1978. — Sept. — Vol. 75, no. 10. — P. 4848—4852.

4. The Conformation of Histone H5: Isolation and Characterisation of the Globular Segment / F. J. AVILES [et al.] // European Journal of Biochemistry. — 1978. — Sept. — Vol. 88, no. 2. — P. 363—371.

5. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome / Z. Peng [et al.] // Cellular and Molecular Life Sciences. — 2013. — Aug. — Vol. 71, no. 8. — P. 1477—1504.

6. Musselman, C. A. Characterization of functional disordered regions within chromatin-associated proteins / C. A. Musselman, T. G. Kutateladze // iScience. — 2021. — Feb. — Vol. 24, no. 2. — P. 102070.

7. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded / D. P. Denning [et al.] // Proceedings of the National Academy of Sciences. — 2003. — Feb. — Vol. 100, no. 5. — P. 2450—2455.

8. Intrinsic structural disorder in cytoskeletal proteins / M. Guharoy [et al.] // Cytoskeleton. — 2013. — June. — Vol. 70, no. 10. — P. 55—57.

9. Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder / V. N. Uversky // Current Opinion in Structural Biology. — 2017. — June. — Vol. 44. — P. 18—30.

10. Uversky Vladimir, N. Intrinsic disorder in proteins associated with neurodegenerative diseases / N. Uversky Vladimir // Frontiers in Bioscience. — 2009. — June. — Vol. 14, no. 1. — P. 5188.

11. Choy, W.-Y. Calculation of ensembles of structures representing the unfolded state of an SH3 domain / W.-Y. Choy, J. D. Forman-Kay // Journal of Molecular Biology. — 2001. — May. — Vol. 308, no. 5. — P. 1011—1032.

12. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from nmr residual dipolar couplings / G. Nodet [et al.] // Journal of the American Chemical Society. — 2009. — Nov. — Vol. 131, no. 49. — P. 17908—17918.

13. Xue, Y. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? / Y. Xue, N. R. Skrynnikov // Journal of the American Chemical Society. — 2011. — Aug. — Vol. 133, no. 37. — P. 14614—14628.

14. Henriques, J. Molecular dynamics simulations of intrinsically disordered proteins: on the accuracy of the TIP4P-D water model and the representativeness of protein disorder models / J. Henriques, M. Skepö // Journal of Chemical Theory and Computation. — 2016. — June. — Vol. 12, no. 7. — P. 3407—3415.

15. Shrestha, U. R. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations / U. R. Shrestha, J. C. Smith, L. Petridis // Communications Biology. — 2021. — Feb. — Vol. 4, no. 1.

16. Best, R. B. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse / R. B. Best, J. Mittal // The Journal of Physical Chemistry B. — 2010. — Nov. — Vol. 114, no. 46. — P. 14916—14923.

17. Nerenberg, P. S. Optimizing protein−solvent force fields to reproduce intrinsic conformational preferences of model peptides / P. S. Nerenberg, T. Head-Gordon // Journal of Chemical Theory and Computation. — 2011. — Mar. — Vol. 7, no. 4. — P. 1220—1230.

18. Water dispersion interactions strongly influence simulated structural properties of disordered protein states / S. Piana [et al.] // The Journal of Physical Chemistry B. — 2015. — Apr. — Vol. 119, no. 16. — P. 5113—5123.

19. Using NMR diffusion data to validate MD models of disordered proteins: Test case of N-terminal tail of histone H4 / O. O. Lebedenko [et al.] // Biophysical Journal. — 2024. — Jan. — Vol. 123, no. 1. — P. 80—100.

20. Conformational and interaction landscape of histone H4 tails in nucleosomes probed by paramagnetic NMR spectroscopy / W. Sun [et al.] // Journal of the American Chemical Society. — 2023. — Nov. — Vol. 145, no. 46. — P. 25478—25485.

21. Aromatic ring flips in differently packed ubiquitin protein crystals from MAS NMR and MD / D. F. Gauto [et al.] // Journal of Structural Biology: X. — 2023. — Dec. — Vol. 7. — P. 1—12.

22. Xue, B. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life / B. Xue, A. K. Dunker, V. N. Uversky // Journal of Biomolecular Structure and Dynamics. — 2012. — June. — Vol. 30, no. 2. — P. 137—149.

23. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome / Z. Peng [et al.] // Cellular and Molecular Life Sciences. — 2013. — Aug. — Vol. 71, no. 8. — P. 1477—1504.

24. Wright, P. E. Intrinsically disordered proteins in cellular signalling and regulation / P. E. Wright, H. J. Dyson // Nature Reviews Molecular Cell Biology. — 2014. — Dec. — Vol. 16, no. 1. — P. 18—29.

25. Context-dependent resistance to proteolysis of intrinsically disordered proteins / M. J. Suskiewicz [et al.] // Protein Science. — 2011. — June. — Vol. 20, no. 8. — P. 1285—1297.

26. Darling, A. L. Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter / A. L. Darling, V. N. Uversky // Frontiers in Genetics. — 2018. — May. — Vol. 9.

27. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins / R. Halfmann [et al.] // Molecular Cell. — 2011. — July. — Vol. 43, no. 1. — P. 72—84.

28. Intrinsic disorder modulates protein self-assembly and aggregation / A. De Simone [et al.] // Proceedings of the National Academy of Sciences. — 2012. — Apr. — Vol. 109, no. 18. — P. 6951—6956.

29. Intrinsic disorder in cell-signaling and cancer-associated Proteins / L. M. Iakoucheva [et al.] // Journal of Molecular Biology. — 2002. — Oct. — Vol. 323, no. 3. — P. 573—584.

30. Targeting intrinsically disordered proteins involved in cancer / P. Santofimia-Castaño [et al.] // Cellular and Molecular Life Sciences. — 2019. — Sept. — Vol. 77, no. 9. — P. 1695—1707.

31. Ji, C. Current status of clinical trials on tau immunotherapies / C. Ji, E. M. Sigurdsson // Drugs Journal. — 2021. — June. — Vol. 81, no. 10. — P. 1135—1152.

32. Immunotherapy for Alzheimer's disease: targeting β-amyloid and beyond / C. Song [et al.] // Translational Neurodegeneration. — 2022. — Mar. — Vol. 11, no. 1.

33. Dyson, H. Coupling of folding and binding for unstructured proteins / H. Dyson, P. E. Wright // Current Opinion in Structural Biology. — 2002. — Feb. — Vol. 12, no. 1. — P. 54—60.

34. Sigalov, A. B. Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form / A. B. Sigalov, A. V. Zhuravleva, V. Y. Orekhov // Biochimie. — 2007. — Mar. — Vol. 89, no. 3. — P. 419—421.

35. Fuxreiter, M. Fuzzy Complexes: polymorphism and structural disorder on protein-protein interactions / M. Fuxreiter, P. Tompa // Biophysical Journal. — 2009. — Feb. — Vol. 96, no. 3. — 319a.

36. Extreme disorder in an ultrahigh-affinity protein complex / A. Borgia [et al.] // Nature. — 2018. — Feb. — Vol. 555, no. 7694. — P. 61—66.

37. Oldfield, C. J. Intrinsically disordered proteins and intrinsically disordered protein regions / C. J. Oldfield, A. K. Dunker // Annual Review of Biochemistry. — 2014. — June. — Vol. 83, no. 1. — P. 553—584.

38. Evidence for Residual Structure in Acid- and Heat-denatured Proteins / K. C. Aune [et al.] // Journal of Biological Chemistry. — 1967. — Oct. — Vol. 242, no. 19. — P. 4486—4489.

39. Hughson, F. M. Structural characterization of a partly folded apomyoglobin intermediate / F. M. Hughson, P. E. Wright, R. L. Baldwin // Science. — 1990. — Sept. — Vol. 249, no. 4976. — P. 1544—1548.

40. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells / W. Borcherds [et al.] // Nature Chemical Biology. — 2014. — Nov. — Vol. 10, no. 12. — P. 1000—1002.

41. Da Vela, S. Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution / S. Da Vela, D. I. Svergun // Current Research in Structural Biology. — 2020. — Vol. 2. — P. 164—170.

42. Smaller capillaries improve the small-angle X-ray scattering signal and sample consumption for biomacromolecular solutions / M. A. Schroer [et al.] // Journal of Synchrotron Radiation. — 2018. — June. — Vol. 25, no. 4. — P. 1113—1122.

43. CHARMM36m: an improved force field for folded and intrinsically disordered proteins / J. Huang [et al.] // Nature Methods. — 2016. — Nov. — Vol. 14, no. 1. — P. 71—73.

44. Robustelli, P. Developing a molecular dynamics force field for both folded and disordered protein states / P. Robustelli, S. Piana, D. E. Shaw // Proceedings of the National Academy of Sciences. — 2018. — May. — Vol. 115, no. 21.

45. Shabane, P. S. General purpose water model can improve atomistic simulations of intrinsically disordered proteins / P. S. Shabane, S. Izadi, A. V. Onufriev // Journal of Chemical Theory and Computation. — 2019. — Mar. — Vol. 15, no. 4. — P. 2620—2634.

46. Accelerating the conformational sampling of intrinsically disordered proteins / T. N. Do [et al.] // Journal of Chemical Theory and Computation. — 2014. — Sept. — Vol. 10, no. 11. — P. 5081—5094.

47. Stanley, N. Kinetic modulation of a disordered protein domain by phosphorylation / N. Stanley, S. Esteban-Martín, G. De Fabritiis // Nature Communications. — 2014. — Oct. — Vol. 5, no. 1.

48. A kinetic ensemble of the Alzheimer's Aβ peptide / T. Löhr [et al.] // Nature Computational Science. — 2021. — Jan. — Vol. 1, no. 1. — P. 71—78.

49. What drives ^{15}N spin relaxation in disordered proteins? combined NMR/MD study of the H4 histone tail / K. Kämpf [et al.] // Biophysical Journal. — 2018. — Dec. — Vol. 115, no. 12. — P. 2348—2367.

50. Histone H4 Tails in nucleosomes: a fuzzy interaction with DNA / S. O. Rabdano [et al.] // Angewandte Chemie International Edition. — 2021. — Feb. — Vol. 60, no. 12. — P. 6480—6487.

51. Salvi, N. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation / N. Salvi, A. Abyzov, M. Blackledge // The Journal of Physical Chemistry Letters. — 2016. — June. — Vol. 7, no. 13. — P. 2483—2489.

52. Dünweg, B. Molecular dynamics simulation of a polymer chain in solution / B. Dünweg, K. Kremer // The Journal of Chemical Physics. — 1993. — Nov. — Vol. 99, no. 9. — P. 6983—6997.

53. Self-diffusion coefficient of the hard-sphere fluid: system size dependence and empirical correlations / D. M. Heyes [et al.] // The Journal of Physical Chemistry B. — 2007. — Jan. — Vol. 111, no. 6. — P. 1455—1464.

54. Raabe, G. Molecular dynamics simulation of the effect of bond flexibility on the transport properties of water / G. Raabe, R. J. Sadus // The Journal of Chemical Physics. — 2012. — Sept. — Vol. 137, no. 10.

55. Aimoli, C. G. Transport properties of carbon dioxide and methane from molecular dynamics simulations / C. G. Aimoli, E. J. Maginn, C. R. A. Abreu // The Journal of Chemical Physics. — 2014. — Oct. — Vol. 141, no. 13.

56. System-size corrections for self-diffusion coefficients calculated from molecular dynamics simulations: The case of CO_2, n-alkanes, and polyethylene glycol) dimethyl ethers / O. A. Moultos [et al.] // The Journal of Chemical Physics. — 2016. — Aug. — Vol. 145, no. 7.

57. Ferrario, V. Simulation of protein diffusion: a sensitive probe of protein–solvent interactions / V. Ferrario, J. Pleiss // Journal of Biomolecular Structure and Dynamics. — 2018. — Apr. — Vol. 37, no. 6. — P. 1534—1544.

58. Characterization of long linear and branched alkanes and alcohols for temperatures up to 573.15 K by surface light scattering and molecular dynamics simulations / T. Klein [et al.] // The Journal of Physical Chemistry B. — 2020. — Apr. — Vol. 124, no. 20. — P. 4146—4163.

59. Yeh, I.-C. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions / I.-C. Yeh, G. Hummer // The Journal of Physical Chemistry B. — 2004. — Sept. — Vol. 108, no. 40. — P. 15873—15879.

60. Dünweg, B. Molecular dynamics algorithms and hydrodynamic screening / B. Dünweg // The Journal of Chemical Physics. — 1993. — Nov. — Vol. 99, no. 9. — P. 6977—6982.

61. Basconi, J. E. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations / J. E. Basconi, M. R. Shirts // Journal of Chemical Theory and Computation. — 2013. — June. — Vol. 9, no. 7. — P. 2887—2899.

62. Hicks, A. Removing Thermostat Distortions of Protein Dynamics in Constant-Temperature Molecular Dynamics Simulations / A. Hicks, M. MacAinsh, H.-X. Zhou // Journal of Chemical Theory and Computation. — 2021. — Aug. — Vol. 17, no. 9. — P. 5920—5932.

63. Bussi, G. Canonical sampling through velocity rescaling / G. Bussi, D. Donadio, M. Parrinello // The Journal of Chemical Physics. — 2007. — Jan. — Vol. 126, no. 1.

64. Hansen, J. C. Structure and function of the core histone N-termini: more than meets the eye / J. C. Hansen, C. Tse, A. P. Wolffe // Biochemistry journal. — 1998. — Dec. — Vol. 37, no. 51. — P. 17637—17641.

65. Self-diffusion coefficient of bulk and confined water: a critical review of classical molecular simulation studies / I. N. Tsimpanogiannis [et al.] // Molecular Simulation. — 2018. — Sept. — Vol. 45, no. 4/5. — P. 425—453.

66. AMBER 2020 / D. Case [et al.]. — University of California, San Francisco, 2020.

67. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB / J. A. Maier [et al.] // Journal of Chemical Theory and Computation. — 2015. — July. — Vol. 11, no. 8. — P. 3696—3713.

68. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew / H. W. Horn [et al.] // The Journal of Chemical Physics. — 2004. — May. — Vol. 120, no. 20. — P. 9665—9678.

69. Izadi, S. Building water models: a different approach / S. Izadi, R. Anandakrishnan, A. V. Onufriev // The Journal of Physical Chemistry Letters. — 2014. — Sept. — Vol. 5, no. 21. — P. 3863—3871.

70. García de la Torre, J. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure / J. García de la Torre, M. L. Huertas, B. Carrasco // Biophysical Journal. — 2000. — Feb. — Vol. 78, no. 2. — P. 719—730.

71. An efficient method for estimating the hydrodynamic radius of disordered protein conformations / M. Nygaard [et al.] // Biophysical Journal. — 2017. — Aug. — Vol. 113, no. 3. — P. 550—557.

72. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins / F. Pesce [et al.] // Biophysical Journal. — 2023. — Jan. — Vol. 122, no. 2. — P. 310—321.

73. Revisiting the use of dioxane as a reference compound for determination of the hydrodynamic radius of proteins by pulsed field gradient NMR spectroscopy / E. E. Tranchant [et al.] // bioRxiv. — 2023. — June.

74. Onufriev, A. V. Water models for biomolecular simulations / A. V. Onufriev, S. Izadi // WIREs Computational Molecular Science. — 2017. — Сент. — Т. 8, № 2.

75. Comparison of simple potential functions for simulating liquid water / W. L. Jorgensen [et al.] // The Journal of Chemical Physics. — 1983. — July. — Vol. 79, no. 2. — P. 926—935.

76. Berendsen, H. J. C. The missing term in effective pair potentials / H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma // The Journal of Physical Chemistry. — 1987. — Nov. — Vol. 91, no. 24. — P. 6269—6271.

77. Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations / D. L. Mobley [et al.] // Journal of Chemical Theory and Computation. — 2009. — Jan. — Vol. 5, no. 2. — P. 350—358.

78. Best, R. B. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association / R. B. Best, W. Zheng, J. Mittal // Journal of Chemical Theory and Computation. — 2014. — Oct. — Vol. 10, no. 11. — P. 5113—5124.

79. Optimizing solute–water van der waals interactions to reproduce solvation free energies / P. S. Nerenberg [et al.] // The Journal of Physical Chemistry B. — 2012. — Apr. — Vol. 116, no. 15. — P. 4524—4534.

80. Abascal, J. L. F. A general purpose model for the condensed phases of water: TIP4P/2005 / J. L. F. Abascal, C. Vega // The Journal of Chemical Physics. — 2005. — Dec. — Vol. 123, no. 23.

81. Improved side-chain torsion potentials for the Amber ff99SB protein force field / K. Lindorff-Larsen [et al.] // Proteins: Structure, Function, and Bioinformatics. — 2010. — Apr. — Vol. 78, no. 8. — P. 1950—1958.

82. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment / S. Rauscher [et al.] // Biophysical Journal. — 2016. — Feb. — Vol. 110, no. 3. — 358a.

83. Miller, M. S. Osmotic pressure simulations of amino acids and peptides highlight potential routes to protein force field parameterization / M. S. Miller,

W. K. Lay, A. H. Elcock // The Journal of Physical Chemistry B. — 2016. — Apr. — Vol. 120, no. 33. — P. 8217—8229.

84. Choice of force field for proteins containing structured and intrinsically disordered regions / V. Zapletal [et al.] // Biophysical Journal. — 2020. — Apr. — Vol. 118, no. 7. — P. 1621—1633.

85. Gil Pineda, L. I. Performance of CHARMM36m with modified water model in simulating intrinsically disordered proteins: a case study / L. I. Gil Pineda, L. N. Milko, Y. He // Biophysics Reports. — 2020. — June. — Vol. 6, no. 2/3. — P. 80—87.

86. Conformational preferences of an intrinsically disordered protein domain: a case study for modern force fields / S. M. Gopal [et al.] // The Journal of Physical Chemistry B. — 2020. — Dec. — Vol. 125, no. 1. — P. 24—35.

87. Abriata, L. A. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization / L. A. Abriata, M. Dal Peraro // Computational and Structural Biotechnology Journal. — 2021. — Vol. 19. — P. 2626—2636.

88. Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment / A. Paul [et al.] // Chemical Science. — 2021. — Vol. 12, no. 19. — P. 6652—6669.

89. Pedersen, K. B. Predicting molecular properties of α-synuclein using force fields for intrinsically disordered proteins / K. B. Pedersen, J. C. Flores-Canales, B. Schiøtt // Proteins: Structure, Function, and Bioinformatics. — 2022. — Aug. — Vol. 91, no. 1. — P. 47—61.

90. Holz, M. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements / M. Holz, S. R. Heil, A. Sacco // Physical Chemistry Chemical Physics. — 2000. — Vol. 2, no. 20. — P. 4740—4742.

91. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature / H. C. Andersen // The Journal of Chemical Physics. — 1980. — Февр. — Т. 72, № 4. — С. 2384—2393.

92. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions / W. G. Hoover // Physical Review A. — 1985. — Mar. — Vol. 31, no. 3. — P. 1695—1697.

93. Acetylation Increases the α-Helical Content of the Histone Tails of the Nucleosome / X. Wang [et al.] // Journal of Biological Chemistry. — 2000. — Sept. — Vol. 275, no. 45. — P. 35013—35020.

94. Potoyan, D. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics / D. A. Potoyan, G. A. Papoian // Journal of the American Chemical Society. — 2011. — Апр. — Т. 133, № 19. — С. 7405—7415.

95. Yang, D. Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation / D. Yang, G. Arya // Phys. Chem. Chem. Phys. — 2011. — Февр. — Т. 13, № 7. — С. 2911—2921.

96. The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects / D. Winogradoff [et al.] // Journal of the American Chemical Society. — 2015. — May. — Vol. 137, no. 19. — P. 6245—6253.

97. Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-Terminal basic patch in the nucleosome / B.-R. Zhou [et al.] // Journal of Molecular Biology. — 2012. — Aug. — Vol. 421, no. 1. — P. 30—37.

98. Histone H3 and H4 N-Terminal Tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy / M. Gao [et al.] // Journal of the American Chemical Society. — 2013. — Oct. — Vol. 135, no. 41. — P. 15278—15281.

99. Shabane, P. S. Significant compaction of H4 histone tail upon charge neutralization by acetylation and its mimics, possible effects on chromatin structure / P. S. Shabane, A. V. Onufriev // Journal of Molecular Biology. — 2021. — Mar. — Vol. 433, no. 6. — P. 166683.

100. Correlating histone acetylation with nucleosome core particle dynamics and function / T. H. Kim [et al.] // Proceedings of the National Academy of Sciences. — 2023. — Apr. — Vol. 120, no. 15.

101. Acetylated histone H4 tail enhances histone H3 tail acetylation by altering their mutual dynamics in the nucleosome / A. Furukawa [et al.] // Proceedings of the National Academy of Sciences. — 2020. — Aug. — Vol. 117, no. 33. — P. 19661—19663.

102. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails / Y. Peng [et al.] // Nature Communications. — 2021. — Sept. — Vol. 12, no. 1.

103. Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry / C. Charlier [et al.] // Journal of the American Chemical Society. — 2013. — Nov. — Vol. 135, no. 49. — P. 18665—18672.

104. Roberts, G. K. NMR Approaches To Understanding Protein Specificity / G. K. Roberts, L.-Y. Lian // Biological NMR Spectroscopy. — Oxford University Press, 02/1997.

105. Determination of an ensemble of structures representing the denatured state of the bovine Acyl-Coenzyme A binding protein / K. Lindorff-Larsen [et al.] // Journal of the American Chemical Society. — 2004. — Feb. — Vol. 126, no. 10. — P. 3291—3299.

106. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins / A. H. Mao [et al.] // Proceedings of the National Academy of Sciences. — 2010. — Apr. — Vol. 107, no. 18. — P. 8183—8188.

107. Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states / J. Lincoff [et al.] // Communications Chemistry. — 2020. — June. — Vol. 3, no. 1.

108. Ortega, A. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models / A. Ortega, D. Amorós, J. García de la Torre // Biophysical Journal. — 2011. — Авг. — Т. 101, № 4. — С. 892—898.

109. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques / W.-Y. Choy [et al.] // Journal of Molecular Biology. — 2002. — Feb. — Vol. 316, no. 1. — P. 101—112.

110. Bernadó, P. A self-consistent description of the conformational behavior of chemically denatured proteins from nmr and small angle scattering / P. Bernadó, M. Blackledge // Biophysical Journal. — 2009. — Сент. — Т. 97, № 10. — С. 2839—2845.

111. Protein dynamics to define and refine disordered protein ensembles / P. M. Naullage [et al.] // The Journal of Physical Chemistry B. — 2022. — Feb. — Vol. 126, no. 9. — P. 1885—1894.

112. Amorós, D. Prediction of hydrodynamic and other solution properties of partially disordered proteins with a simple, coarse-grained model / D. Amorós, A. Ortega, J. García de la Torre // Journal of Chemical Theory and Computation. — 2013. — Feb. — Vol. 9, no. 3. — P. 1678—1685.

113. García de la Torre, J. Hydrodynamic properties of biomacromolecules and macromolecular complexes: concepts and methods. a tutorial mini-review / J. García de la Torre, J. Hernández Cifre // Journal of Molecular Biology. — 2020. — Apr. — Vol. 432, no. 9. — P. 2930—2948.

114. Braga, C. Computer simulation of the role of torsional flexibility on mass and momentum transport for a series of linear alkanes / C. Braga, K. P. Travis // The Journal of Chemical Physics. — 2012. — Aug. — Vol. 137, no. 6.

115. Bulacu, M. Effect of bending and torsion rigidity on self-diffusion in polymer melts: A molecular-dynamics study / M. Bulacu, E. van der Giessen // The Journal of Chemical Physics. — 2005. — Sept. — Vol. 123, no. 11.

116. The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule / E. Brookes [et al.] // European Biophysics Journal. — 2009. — Feb. — Vol. 39, no. 3. — P. 423—435.

117. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables / V. Ozenne [et al.] // Bioinformatics. — 2012. — May. — Vol. 28, no. 11. — P. 1463—1470.

118. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties / G. Tesei [et al.] // Proceedings of the National Academy of Sciences. — 2021. — Oct. — Vol. 118, no. 44.

119. Kirkwood, J. G. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution / J. G. Kirkwood, J. Riseman // The Journal of Chemical Physics. — 1948. — June. — Vol. 16, no. 6. — P. 565—573.

120. Fleming, P. J. Revisiting macromolecular hydration with HullRadSAS / P. J. Fleming, J. J. Correia, K. G. Fleming // European Biophysics Journal. — 2023. — Jan. — Vol. 52, no. 4/5. — P. 215—224.

121. AMBER 2023 / D. Case [et al.]. — University of California, San Francisco, 2023.

122. Miyamoto, S. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models / S. Miyamoto, P. A. Kollman // Journal of Computational Chemistry. — 1992. — Oct. — Vol. 13, no. 8. — P. 952—962.

123. Jung, J. Optimal temperature and pressure evaluations in molecular dynamics simulations with a large time step / J. Jung, C. Kobayashi, Y. Sugita // Biophysical Journal. — 2019. — Feb. — Vol. 116, no. 3. — 141a.

124. Statistical coil model of the unfolded state: Resolving the reconciliation problem / A. K. Jha [et al.] // Proceedings of the National Academy of Sciences. — 2005. — Aug. — Vol. 102, no. 37. — P. 13099—13104.

125. Krivov, G. G. Improved prediction of protein side-chain conformations with SCWRL4 / G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack // Proteins: Structure, Function, and Bioinformatics. — 2009. — July. — Vol. 77, no. 4. — P. 778—795.

126. Nguyen, H. Improved generalized born solvent model parameters for protein simulations / H. Nguyen, D. R. Roe, C. Simmerling // Chemical Theory and computation. — 2013. — Mar. — Vol. 9, no. 4. — P. 2020—2034.

127. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKaPredictions / M. H. M. Olsson [et al.] // Journal of Chemical Theory and Computation. — 2011. — Jan. — Vol. 7, no. 2. — P. 525—537.

128. Joung, I. S. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations / I. S. Joung, T. E. Cheatham // The Journal of Physical Chemistry B. — 2008. — July. — Vol. 112, no. 30. — P. 9020—9041.

129. Molecular dynamics with coupling to an external bath / H. J. C. Berendsen [et al.] // The Journal of Chemical Physics. — 1984. — Oct. — Vol. 81, no. 8. — P. 3684—3690.

130. Ryckaert, J.-P. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes / J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen // Journal of Computational Physics. — 1977. — Mar. — Vol. 23, no. 3. — P. 327—341.

131. Effects of molecular dynamics thermostats on descriptions of chemical nonequilibrium / A. J. Page [et al.] // Journal of Chemical Theory and Computation. — 2012. — Oct. — Vol. 8, no. 11. — P. 4019—4028.

132. Frenkel, D. Understanding molecular simulation: from algorithms to applications / D. Frenkel, B. Smit, M. A. Ratner // Physics Today. — 1997. — July. — Vol. 50, no. 7. — P. 66—66.

133. Vijay-kumar, S. Structure of ubiquitin refined at 1.8 Å resolution / S. Vijay-kumar, C. E. Bugg, W. J. Cook // Journal of Molecular Biology. — 1987. — Apr. — Vol. 194, no. 3. — P. 531—544.

134. Long-time-step molecular dynamics through hydrogen mass repartitioning / C. W. Hopkins [et al.] // Journal of Chemical Theory and Computation. — 2015. — Mar. — Vol. 11, no. 4. — P. 1864—1874.

135. Bülow, S. von. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure / S. von Bülow, J. T. Bullerjahn, G. Hummer // The Journal of Chemical Physics. — 2020. — July. — Vol. 153, no. 2.

136. Haile, J. M. Molecular dynamics simulation / J. M. Haile. — Nashville, TN: John Wiley Sons, 02/1997. — (Wiley professional paperback series).

137. Qian, H. Single particle tracking. analysis of diffusion and flow in two-dimensional systems / H. Qian, M. Sheetz, E. Elson // Biophysical Journal. — 1991. — Oct. — Vol. 60, no. 4. — P. 910—921.

138. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium / X. Michalet // Physical Review E. — 2010. — Oct. — Vol. 82, no. 4.

139. Fliege, J. The distribution of points on the sphere and corresponding cubature formulae / J. Fliege // IMA Journal of Numerical Analysis. — 1999. — Apr. — Vol. 19, no. 2. — P. 317—334.

140. Wong, V. Evaluating rotational diffusion from protein md simulations / V. Wong, D. A. Case // The Journal of Physical Chemistry B. — 2007. — Dec. — Vol. 112, no. 19. — P. 6013—6024.

141. Numerical recipes in C: the art of scientific computing / William H. Press [et al.]. — 2nd ed. — Cambridge, England: Cambridge University Press, 10/1992.

142. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation / N. Tjandra [et al.] // Journal of the American Chemical Society. — 1995. — Dec. — Vol. 117, no. 50. — P. 12562—12566.

143. Linke, M. Rotational diffusion depends on box size in molecular dynamics simulations / M. Linke, J. Köfinger, G. Hummer // The Journal of Physical Chemistry Letters. — 2018. — May. — Vol. 9, no. 11. — P. 2874—2878.

144. SciPy 1.0: fundamental algorithms for scientific computing in Python / P. Virtanen [et al.] // Nature Methods. — 2020. — Feb. — Vol. 17, no. 3. — P. 261—272.

145. Palmer, A. G. I. NMR characterization of the dynamics of biomacromolecules / A. G. I. Palmer // ChemInform journal. — 2004. — Oct. — Vol. 35, no. 45.

146. Palmer III, A. G. NMR probes of molecular dynamics: overview and comparison with other techniques / A. G. Palmer III // Annual Review of Biophysics and Biomolecular Structure. — 2001. — June. — Vol. 30, no. 1. — P. 129—155.

147. Binary mutual diffusion coefficients of aqueous cyclic ethers at 25 °c. tetrahydrofuran, 1,3-dioxolane, 1,4-dioxane, 1,3-dioxane, tetrahydropyran, and trioxane / D. G. Leaist [et al.] // Journal of Chemical amp; Engineering Data. — 2000. — Sept. — Vol. 45, no. 5. — P. 815—818.

148. Clisby, N. High-precision estimate of the hydrodynamic radius for self-avoiding walks / N. Clisby, B. Dünweg // Physical Review E. — 2016. — Nov. — Vol. 94, no. 5.

149. Lazar, G. A. De novo design of the hydrophobic core of ubiquitin / G. A. Lazar, J. R. Desjarlais, T. M. Handel // Protein Science. — 1997. — June. — Vol. 6, no. 6. — P. 1167—1178.

150. Noncovalent dimerization of ubiquitin / Z. Liu [et al.] // Angewandte Chemie. — 2011. — Nov. — Vol. 124, no. 2. — P. 484—487.

151. Price, W. S. Diffusion and its measurement / W. S. Price. — Cambridge: Cambridge University Press, 07/2009. — P. 1—68.

152. Jerschow, A. Suppression of convection artifacts in stimulated-echo diffusion experiments. double-stimulated-echo experiments / A. Jerschow, N. Müller // Journal of Magnetic Resonance. — 1997. — Apr. — Vol. 125, no. 2. — P. 372—375.

153. Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity / V. Sklenar [et al.] // Journal of Magnetic Resonance, Series A. — 1993. — Apr. — Vol. 102, no. 2. — P. 241—245.

154. The role of rotational motion in diffusion nmr experiments on supramolecular assemblies: application to Sup35NM fibrils / B. B. Kharkov [et al.] // Angewandte Chemie. — 2021. — June. — Vol. 133, no. 28. — P. 15573—15579.

155. Stejskal, E. O. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient / E. O. Stejskal, J. E. Tanner // The Journal of Chemical Physics. — 1965. — Jan. — Vol. 42, no. 1. — P. 288—292.

156. Hermann, M. R. Saxs-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy / M. R. Hermann, J. S. Hub // Journal of Chemical Theory and Computation. — 2019. — Aug. — Vol. 15, no. 9. — P. 5103—5115.

157. // NMR studies of translational motion. — Cambridge University Press, 07/2009. — P. 296—307.

158. Boggs, P. T. The computation and use of the asymptotic covariance matrix for measurement error models / P. T. Boggs, J. R. Donaldson. — 1989.

159. Whitlock, J. Folding of DNA by histones which lack their NH2-terminal regions. / J. Whitlock, A. Stein // Journal of Biological Chemistry. — 1978. — June. — Vol. 253, no. 11. — P. 3857—3861.

160. Cary Peter D. Moss, T. High-resolution proton-magnetic-resonance studies of chromatin core particles / T. Cary Peter D. Moss, E. M. Bradbury // European Journal of Biochemistry. — 1978. — Sept. — Vol. 89, no. 2. — P. 475—482.

161. Crystal structure of the nucleosome core particle at 2.8 Å resolution / K. Luger [et al.] // Nature journal. — 1997. — Sept. — Vol. 389, no. 6648. — P. 251—260.

162. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution / C. A. Davey [et al.] // Journal of Molecular Biology. — 2002. — June. — Vol. 319, no. 5. — P. 1097—1113.

163. Jenuwein, T. Translating the histone code / T. Jenuwein, C. D. Allis // Science. — 2001. — Aug. — Vol. 293, no. 5532. — P. 1074—1080.

164. Bannister, A. J. Regulation of chromatin by histone modifications / A. J. Bannister, T. Kouzarides // Cell Research. — 2011. — Feb. — Vol. 21, no. 3. — P. 381—395.

165. Strahl, B. D. The language of covalent histone modifications / B. D. Strahl, C. D. Allis // Nature. — 2000. — Jan. — Vol. 403, no. 6765. — P. 41—45.

166. The role of the nucleosome acidic patch in modulating higher order chromatin structure / A. A. Kalashnikova [et al.] // Journal of The Royal Society Interface. — 2013. — May. — Vol. 10, no. 82. — P. 20121022.

167. Structure and dynamics in the nucleosome revealed by solid-state NMR / X. Shi [et al.] // Angewandte Chemie International Edition. — 2018. — July. — Vol. 57, no. 31. — P. 9734—9738.

168. Ghoneim, M. Histone tail conformations: a fuzzy affair with DNA / M. Ghoneim, H. A. Fuchs, C. A. Musselman // Trends in Biochemical Sciences. — 2021. — July. — Vol. 46, no. 7. — P. 564—578.

169. Histone H3 and H4 tails play an important role in nucleosome phase separation / E. F. Hammonds [et al.] // Biophysical Chemistry. — 2022. — Apr. — Vol. 283. — P. 106767.

170. Modulations of DNA contacts by linker histones and post-translational modifications determine the mobility and modifiability of nucleosomal H3 tails / A. Stützer [et al.] // Molecular Cell. — 2016. — Jan. — Vol. 61, no. 2. — P. 247—259.

171. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome / E. A. Morrison [et al.] // eLife. — 2018. — Apr. — Vol. 7.

172. Conformational dynamics of histone H3 tails in chromatin / M. Zandian [et al.] // The Journal of Physical Chemistry Letters. — 2021. — June. — Vol. 12, no. 26. — P. 6174—6181.

173. Nucleosome composition regulates the histone H3 tail conformational ensemble and accessibility / E. A. Morrison [et al.] // Nucleic Acids Research. — 2021. — Apr. — Vol. 49, no. 8. — P. 4750—4767.

174. Dynamic networks observed in the nucleosome core particles couple the histone globular domains with DNA / X. Shi [et al.] // Communications Biology. — 2020. — Oct. — Vol. 3, no. 1.

175. Histone H3 core domain in chromatin with different DNA linker lengths studied by ^{1}H-Detected solid-state NMR spectroscopy / S. T. Smrt [et al.] // Frontiers in Molecular Biosciences. — 2023. — Jan. — Vol. 9.

176. Site-specific studies of nucleosome interactions by solid-state NMR spectroscopy / S. Xiang [et al.] // Angewandte Chemie International Edition. — 2018. — Mar. — Vol. 57, no. 17. — P. 4571—4575.

177. The N-terminal tails of histones H2A and H2B adopt two distinct conformations in the nucleosome with contact and reduced contact to DNA / H. Ohtomo [et al.] // Journal of Molecular Biology. — 2021. — July. — Vol. 433, no. 15. — P. 167110.

178. Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes / M. Shoaib [et al.] // Nature Communications. — 2021. — Aug. — Vol. 12, no. 1.

179. Jennings, C. E. Arginine anchor points govern H3 tail dynamics / C. E. Jennings, C. J. Zoss, E. A. Morrison // Frontiers in Molecular Biosciences. — 2023. — May. — Vol. 10.

180. Musselman, C. A. Visualizing conformational ensembles of the nucleosome by NMR / C. A. Musselman, T. G. Kutateladze // ACS Chemical Biology. — 2022. — Feb. — Vol. 17, no. 3. — P. 495—502.

181. Battiste, J. L. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited Nuclear Overhauser Effect data / J. L. Battiste, G. Wagner // Biochemistry journal. — 2000. — Apr. — Vol. 39, no. 18. — P. 5355—5365.

182. Protein global fold determination using site-directed spin and isotope labeling / V. Gaponenko [et al.] // Protein Science. — 2000. — Jan. — Vol. 9, no. 2. — P. 302—309.

183. Iwahara, J. Ensemble approach for NMR structure refinement against ^{1}H Paramagnetic Relaxation Enhancement data arising from a flexible paramagnetic group attached to a macromolecule / J. Iwahara, C. D. Schwieters, G. M. Clore // Journal of the American Chemical Society. — 2004. — Apr. — Vol. 126, no. 18. — P. 5879—5896.

184. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR / H. Kato [et al.] // Proceedings of the National Academy of Sciences. — 2011. — July. — Vol. 108, no. 30. — P. 12283—12288.

185. Structural insights into the histone H1-nucleosome complex / B.-R. Zhou [et al.] // Proceedings of the National Academy of Sciences. — 2013. — Sept. — Vol. 110, no. 48. — P. 19390—19395.

186. A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome / B.-R. Zhou [et al.] // Journal of Molecular Biology. — 2016. — Oct. — Vol. 428, no. 20. — P. 3948—3959.

187. Solomon, I. Nuclear magnetic interactions in the HF molecule / I. Solomon, N. Bloembergen // The Journal of Chemical Physics. — 1956. — Aug. — Vol. 25, no. 2. — P. 261—266.

188. Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and ^{13}Cα nuclear spin relaxation / L. K. Lee [et al.] // J. Biomol. NMR. — 1997. — Apr. — Vol. 9, no. 419. — P. 287—298.

189. Iwahara, J. Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long, variable-length vectors using NMR Paramagnetic Relaxation Enhancement / J. Iwahara, G. M. Clore // Journal of the American Chemical Society. — 2010. — Aug. — Vol. 132, no. 38. — P. 13346—13356.

190. Mapping long-range contacts in a highly unfolded protein / M. A. Lietzow [et al.] // Journal of Molecular Biology. — 2002. — Sept. — Vol. 322, no. 4. — P. 655—662.

191. Teilum, K. Transient structure formation in unfolded acyl-coenzyme a-binding protein observed by site-directed spin labelling / K. Teilum, B. B. Kragelund, F. M. Poulsen // Journal of Molecular Biology. — 2002. — Nov. — Vol. 324, no. 2. — P. 349—357.

192. Characterization of conformational and dynamic properties of natively unfolded human and mouse α-synuclein ensembles by NMR: implication for aggregation / K.-P. Wu [et al.] // Journal of Molecular Biology. — 2008. — May. — Vol. 378, no. 5. — P. 1104—1115.

193. Gillespie, J. R. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation enhancement by nitroxide spin labels / J. R. Gillespie, D. Shortle // Journal of Molecular Biology. — 1997. — Apr. — Vol. 268, no. 1. — P. 158—169.

194. Huang, J.-r. Ensemble Calculations of Unstructured Proteins Constrained by RDC and PRE Data: A Case Study of Urea-Denatured Ubiquitin / J.-r. Huang, S. Grzesiek // Journal of the American Chemical Society. — 2009. — Dec. — Vol. 132, no. 2. — P. 694—705.

195. NMR characterization of long-range order in intrinsically disordered proteins / L. Salmon [et al.] // Journal of the American Chemical Society. — 2010. — May. — Vol. 132, no. 24. — P. 8407—8418.

196. Dedmon, M. M. Mapping long-range interactions in α-synuclein using spin-label nmr and ensemble molecular dynamics simulations / M. M. Dedmon, Lindorff-Larsen, others. // Journal of the American Chemical Society. — 2004. — Dec. — Vol. 127, no. 2. — P. 476—477.

197. Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements / J. R. Allison [et al.] // Journal of the American Chemical Society. — 2009. — Dec. — Vol. 131, no. 51. — P. 18314—18326.

198. Ganguly, D. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states / D. Ganguly, J. Chen // Journal of Molecular Biology. — 2009. — July. — Vol. 390, no. 3. — P. 467—477.

199. Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure / J. A. Marsh [et al.] // Journal of Molecular Biology. — 2007. — Apr. — Vol. 367, no. 5. — P. 1494—1510.

200. Modulation of correlated segment fluctuations in IDPs upon complex formation as an allosteric regulatory mechanism / A. Beier [et al.] // Journal of Molecular Biology. — 2018. — Aug. — Vol. 430, no. 16. — P. 2439—2452.

201. Paramagnetic relaxation enhancements in unfolded proteins: Theory and application to drkN SH3 domain / Y. Xue [et al.] // Protein Science. — 2009. — June. — Vol. 18, no. 7. — P. 1401—1424.

202. Accelerating molecular modeling applications with graphics processors / J. E. Stone [et al.] // Journal of Computational Chemistry. — 2007. — Sept. — Vol. 28, no. 16. — P. 2618—2640.

203. Huang, J. Force field development and simulations of intrinsically disordered proteins / J. Huang, A. D. MacKerell // Current Opinion in Structural Biology. — 2018. — Feb. — Vol. 48. — P. 40—48.

204. Recent Force Field Strategies for intrinsically disordered proteins / J. Mu [et al.] // Journal of Chemical Information and Modeling. — 2021. — Feb. — Vol. 61, no. 3. — P. 1037—1047.

205. Lowary, P. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning / P. Lowary, J. Widom // Journal of Molecular Biology. — 1998. — Feb. — Vol. 276, no. 1. — P. 19—42.

206. Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations / C. Altenbach [et al.] // Biochemistry journal. — 2001. — Nov. — Vol. 40, no. 51. — P. 15471—15482.

207. Recent advances in site-directed spin labeling of proteins / W. L. Hubbell [et al.] // Current Opinion in Structural Biology. — 1998. — Oct. — Vol. 8, no. 5. — P. 649—656.

208. Izadi, S. Building water models: a different approach / S. Izadi, R. Anandakrishnan, A. V. Onufriev // The Journal of Physical Chemistry Letters. — 2014. — Sept. — Vol. 5, no. 21. — P. 3863—3871.

209. Vasudevan, D. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence / D. Vasudevan, E. Y. Chua, C. A. Davey // Journal of Molecular Biology. — 2010. — Oct. — Vol. 403, no. 1. — P. 1—10.

210. Reconstitution of nucleosome core particles from recombinant histones and DNA / P. N. Dyer [et al.] // Methods in Enzymology. — Elsevier, 2003. — P. 23—44.

211. Luger, K. Preparation of nucleosome core particle from recombinant histones / K. Luger, T. J. Rechsteiner, T. J. Richmond // Chromatin. — Elsevier, 1999. — P. 3—19.

212. Structural and dynamic origins of ESR lineshapes in spin-labeled GB1 domain: the insights from spin dynamics simulations based on long MD trajectories / S. A. Izmailov [et al.] // Scientific Reports. — 2020. — Jan. — Vol. 10, no. 1.

213. Humphrey, W. VMD: visual molecular dynamics / W. Humphrey, A. Dalke, K. Schulten // Journal of Molecular Graphics. — 1996. — Feb. — Vol. 14, no. 1. — P. 33—38.

214. IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states / J. M. C. Teixeira [et al.] // The Journal of Physical Chemistry A. — 2022. — Aug. — Vol. 126, no. 35. — P. 5985—6003.

215. Webb, B. Comparative protein structure modeling using MODELLER / B. Webb, A. Sali // Current Protocols in Bioinformatics. — 2016. — June. — Vol. 54, no. 1.

216. Aromatic clusters in protein-protein and protein-drug complexes / E. Lanzarotti [et al.] // Journal of Cheminformatics. — 2020. — Vol. 12, no. 1. — P. 30.

217. Wuthrich, K. Internal motion in globular proteins / K. Wuthrich, G. Wagner // Trends Biochem. Sci. — 1977. — Vol. 3, no. 10. — P. 227—230.

218. Campbell, I. D. Proton magnetic resonance studies of the tyrosine residues of hen lysozyme-assignment and detection of conformational mobility / I. D. Campbell, C. M. Dobson, R. J. P. Williams // Proc. R. Soc. B. — 1975. — Vol. 189, no. 1097. — P. 503—509.

219. Wagner, G. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. ^{1}H NMR studies. / G. Wagner, a. DeMarco, K. Wüthrich // Biophys. Struct. Mech. — 1976. — Vol. 2, no. 2. — P. 139—158.

220. Protein dynamics by solid-state NMR: aromatic rings of the coat protein in fd bacteriophage / C. Gall [et al.] // Proc. Natl. Acad. Sci. — 1982. — Vol. 79, no. 1. — P. 101—105.

221. Hu, F. Mechanisms of proton conduction and gating in influenza M2 proton channels from solid-state NMR / F. Hu, W. Luo, M. Hong // Science. — 2010. — Vol. 330, no. 6003. — P. 505—508.

222. Dynamics of aromatic side chains in the active site of FKBP12 / U. Weininger [et al.] // Biochemistry journal. — 2017. — Vol. 56, no. 1. — P. 334—343.

223. Slow aromatic ring flips detected despite near-degenerate NMR frequencies of the exchanging nuclei / U. Weininger [et al.] // Journal of Physical Chemistry. — 2013. — Vol. 117, no. 31. — P. 9241—9247.

224. Weininger, U. Ring flips revisited: ^{13}C relaxation dispersion measurements of aromatic side chain dynamics and activation barriers in basic pancreatic trypsin inhibitor / U. Weininger, K. Modig, M. Akke // Biochemistry journal. — 2014. — Vol. 53, no. 28. — P. 4519—4525.

225. Raum, H. N. Conformational exchange of aromatic side chains by ^{1}H CPMG relaxation dispersion / H. N. Raum, M. Dreydoppel, U. Weininger // J. Biomol. NMR. — 2018. — Vol. 72, no. 1. — P. 105—114.

226. Bremi, T. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: application to phenylalanines in antamanide / T. Bremi, R. Brüschweiler, R. R. Ernst // Journal of the American Chemical Society. — 1997. — Vol. 119, no. 18. — P. 4272—4284.

227. Transition-state compressibility and activation volume of transient protein conformational fluctuations / M. Dreydoppel [et al.] // JACS Au. — 2021. — Vol. 1, no. 6. — P. 833—842.

228. Infrequent cavity-forming fluctuations in HPr from Staphylococcus carnosus revealed by pressure-and temperature-dependent tyrosine ring flips / M. Hattori [et al.] // Protein Sci. — 2004. — Vol. 13, no. 12. — P. 3104—3114.

229. A sharp thermal transition of fast aromatic-ring dynamics in ubiquitin. / V. Kasinath [et al.] // Angew. Chem. Int. Ed. — 2015. — Vol. 54, no. 1. — P. 102—107.

230. Visualizing protein breathing motions associated with aromatic ring flipping / L. Marino P'erez [et al.] // Nature. — 2022. — Vol. 602, no. 7898. — P. 695—700.

231. Kasinath, V. A ^{13}C labeling strategy reveals a range of aromatic side chain motion in calmodulin. / V. Kasinath, K. G. Valentine, A. J. Wand // Journal of the American Chemical Society. — 2013. — Vol. 135, no. 26. — P. 9560—9563.

232. Gall, C. Phenylalanine ring dynamics by solid-state deuterium NMR / C. Gall, J. DiVerdi, S. Opella // Journal of the American Chemical Society. — 1981. — Vol. 103, no. 17. — P. 5039—5043.

233. Dynamics of hydrophobic core phenylalanine residues probed by solid-state deuteron NMR / L. Vugmeyster [et al.] // Journal of Physical Chemistry. — 2015. — Vol. 119, no. 47. — P. 14892—14904.

234. Solvent-driven dynamical crossover in the phenylalanine side-chain from the hydrophobic core of amyloid fibrils detected by ^{2}H NMR relaxation / L. Vugmeyster [et al.] // Journal of Physical Chemistry. — 2017. — Vol. 121, no. 30. — P. 7267—7275.

235. Bonaccorsi, M. Protein structural dynamics by magic-angle spinning NMR / M. Bonaccorsi, T. Le Marchand, G. Pintacuda // Curr. Opin. Struct. Biol. — 2021. — T. 70. — C. 34—43.

236. Lewandowski, J. R. Advances in solid-state relaxation methodology for probing site-specific protein dynamics / J. R. Lewandowski // Acc. Chem. Res. — Department of Chemistry, University of Warwick , Coventry CV4 7AL, U.K., 2013. — Vol. 46, no. 9. — P. 2018—2027.

237. Schanda, P. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules / P. Schanda, M. Ernst // Prog. Nucl. Magn. Reson. Spectrosc. — 2016. — T. 96. — C. 1—46.

238. Krushelnitsky, A. Solid-state NMR and protein dynamics / A. Krushelnitsky, D. Reichert // Prog. Nucl. Magn. Reson. Spectr. — 2005. — Vol. 47, no. 1/2. — P. 1—25.

239. Probing structure and dynamics of protein assemblies by magic angle spinning NMR spectroscopy / S. Yan [et al.] // Acc. Chem. Res. — 2013. — Vol. 46, no. 9. — P. 2047—2058.

240. Reif, B. Deuterated peptides and proteins: structure and dynamics studies by MAS solid-state NMR / B. Reif // Protein NMR Techniques. — Springer, 2012. — C. 279—301.

241. Watt, E. D. Recent advances in solid-state nuclear magnetic resonance techniques to quantify biomolecular dynamics / E. D. Watt, C. M. Rienstra // Anal. Chem. — 2014. — Vol. 86, no. 1. — P. 58—64.

242. Krushelnitsky, A. Solid-state NMR approaches to internal dynamics of proteins: from picoseconds to microseconds and seconds / A. Krushelnitsky, D. Reichert, K. Saalwachter // Acc. Chem. Res. — 2013. — Vol. 46, no. 9. — P. 2028—2036.

243. Aromatic ring dynamics, thermal activation, and transient conformations of a 468 kDa enzyme by specific ^{1}H–^{13}C labeling and fast magic-angle spinning NMR / D. F. Gauto [et al.] // Journal of the American Chemical Society. — 2019. — Vol. 141, no. 28. — P. 11183—11195.

244. Slow conformational exchange and overall rocking motion in ubiquitin protein crystals / V. Kurauskas [et al.] // Nat. Commun. — 2017. — Vol. 8, no. 1. — P. 1—12.

245. Probing transient conformational states of proteins by solid-state $R_{1\rho}$ Relaxation-Dispersion NMR Spectroscopy / P. Ma [et al.] // Angew. Chem. Int. Ed. — 2014. — Vol. 53, no. 17. — P. 4312—4317.

246. Solid-state NMR Measurements of asymmetric dipolar couplings provide insight into protein Side-Chain Motion / P. Schanda [et al.] // Angew. Chem. Int. Ed. — 2011. — Vol. 50, no. 46. — P. 11005—11009.

247. Smith, A. A. Because the light is better here: correlation-time analysis by NMR spectroscopy / A. A. Smith, M. Ernst, B. H. Meier // Angew. Chem. Int. Ed. — 2017. — Vol. 129, no. 44. — P. 13778—13783.

248. Zumpfe, K. Model-free or not? / K. Zumpfe, A. A. Smith // Front. Mol. Biosci. — 2021. — T. 8. — C. 727553.

249. Observing the overall rocking motion of a protein in a crystal / P. Ma [et al.] // Nat. Commun. — 2015. — Vol. 6, no. 1. — P. 1—10.

250. ^{1}H-detected biomolecular NMR under fast magic-angle spinning / T. Le Marchand [et al.] // Chem. Rev. — 2022. — Vol. 122. — P. 9943—10018.

251. High-resolution proton-detected NMR of proteins at very fast MAS / L. B. Andreas [et al.] // J. Magn. Reson. — 2015. — Vol. 253. — P. 36—49.

252. Optimal degree of protonation for ^{1}H detection of aliphatic sites in randomly deuterated proteins as a function of the MAS frequency / S. Asami [et al.] // J. Biomol. NMR. — 2012. — Vol. 54, no. 2. — P. 155—168.

253. Agarwal, V. Residual methyl protonation in perdeuterated proteins for multi-dimensional correlation experiments in MAS solid-state NMR spectroscopy / V. Agarwal, B. Reif // J. Magn. Reson. — 2008. — Vol. 194, no. 1. — P. 16—24.

254. Ruschak, A. M. Methyl groups as probes of supra-molecular structure, dynamics and function / A. M. Ruschak, L. E. Kay // J. Biomol. NMR. — 2010. — Vol. 46, no. 1. — P. 75—87.

255. Methyl-specific isotopic labeling: A molecular tool box for solution NMR studies of large proteins / R. Kerfah [et al.] // Curr. Opin. Struct. Biol. — 2015. — Vol. 32. — P. 113—122.

256. Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: A similarity revealed / V. Agarwal [et al.] // Journal of the American Chemical Society. — 2008. — Vol. 130, no. 49. — P. 16611—16621.

257. α-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. / R. J. Lichtenecker [et al.] // J. Biomol. NMR. — Institute of Organic Chemistry, University of Vienna, Währingerstr. 38, 1090, Vienna, Austria, roman.lichtenecker@univie.ac.at., 2013. — Vol. 57, no. 4. — P. 327—31.

258. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase / G. Cornilescu [et al.] // Journal of the American Chemical Society. — 1998. — Vol. 120, no. 27. — P. 6836—6837.

259. Babu, C. R. Validation of protein structure from preparations of encapsulated proteins dissolved in low viscosity fluids / C. R. Babu, P. F. Flynn, A. J. Wand // Journal of the American Chemical Society. — 2001. — Vol. 123, no. 11. — P. 2691—2692.

260. Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system / T. Ikeya [et al.] // J. Biomol. NMR. — 2009. — Vol. 44, no. 4. — P. 261—272.

261. High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells / T. Tanaka [et al.] // Angew. Chem. Int. Ed. — 2019. — Vol. 131, no. 22. — P. 7362—7366.

262. Structural heterogeneity in microcrystalline ubiquitin studied by solid-state NMR / H. K. Fasshuber [et al.] // Prot. Sci. — 2015. — Vol. 24, no. 5. — P. 592—598.

263. SHIFTX2: significantly improved protein chemical shift prediction / B. Han [et al.] // J. Biomol. NMR. — 2011. — Vol. 50, no. 1. — P. 43—57.

264. Haigh, C. Ring current theories in nuclear magnetic resonance / C. Haigh, R. Mallion // Prog. Nucl. Magn. Reson. Spectrosc. — 1979. — Vol. 13, no. 4. — P. 303—344.

265. Case, D. A. Calibration of ring-current effects in proteins and nucleic acids / D. A. Case // Journal of Biomolecular NMR. — 1995. — Vol. 6, no. 4. — P. 341—346.

266. Hess, S. Tensors for physics / S. Hess. — Springer, 2015.

267. Gullion, T. Detection of weak heteronuclear dipolar coupling by rotational-echo double-resonance nuclear magnetic resonance / T. Gullion, J. Schaefer // Advances in Magnetic and Optical Resonance. T. 13. — Elsevier, 1989. — C. 57—83.

268. Double-quantum homonuclear rotary resonance - efficient dipolar recovery in magic angle spinning nuclear magnetic resonance / N. C. Nielsen [et al.] // J. Chem. Phys. — 1994. — Vol. 101, no. 3. — P. 1805—1812.

269. Mechanistic insights into microsecond time-scale motion of solid proteins using complementary ^{15}N and ^{1}H relaxation dispersion techniques / P. Rovó [et al.] // Journal of the American Chemical Society. — 2019. — Vol. 141, no. 2. — P. 858—869.

270. Protein conformational dynamics studied by ^{15}N and ^{1}H $R_{1\rho}$ relaxation dispersion: Application to wild-type and G53A ubiquitin crystals / D. F. Gauto [et al.] // Solid State Nucl. Magn. Reson. — 2017. — Oct. — Vol. 87. — P. 86—95.

271. VanderHart, D. ^{13}C NMR rotating frame relaxation in a solid with strongly coupled protons: Polyethylene / D. VanderHart, A. Garroway // J. Chem. Phys. — 1979. — Vol. 71, no. 7. — P. 2773—2787.

272. Spin–spin and spin-lattice contributions to the rotating frame relaxation of ^{13}C in L-alanine / K. Akasaka [et al.] // J. Chem. Phys. — 1983. — Vol. 78, no. 6. — P. 3567—3572.

273. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB / J. A. Maier [et al.] // J. Chem. Theory Comput. — 2015. — Vol. 11, no. 8. — P. 3696—3713.

274. ssNMRlib: a comprehensive library and tool box for acquisition of solid-state nuclear magnetic resonance experiments on Bruker spectrometers / A. Vallet [et al.] // Magn. Reson. — 2020. — Vol. 1, no. 2. — P. 331—345.

275. Schanda, P. Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR Spectroscopy / P. Schanda, B. H. Meier, M. Ernst // Journal of the American Chemical Society. — 2010. — Vol. 132, no. 45. — P. 15957—15967.

276. Cross-correlated relaxation of dipolar coupling and chemical-shift anisotropy in magic-angle spinning $R_{1\rho}$ NMR measurements: application to protein backbone dynamics measurements / V. Kurauskas [et al.] // Journal of Physical Chemistry. — 2016. — Sept. — Vol. 120, no. 34. — P. 8905—8913.

277. Computer simulations in magnetic resonance. An object-oriented programming approach / S. Smith [et al.] // J. Magn. Reson. — 1994. — Vol. 106. — P. 75—105.

278. Kurbanov, R. The nuclear magnetic resonance relaxation data analysis in solids: general R_1/ $R_{1\rho}$ equations and the model-free approach / R. Kurbanov, T. Zinkevich, A. Krushelnitsky // J. Chem. Phys. — 2011. — Vol. 135, no. 18. — P. 184104.

279. Analysis of the backbone dynamics of interleukin-1b using two-dimensional inverse detected heteronuclear ^{15}N–^{1}H NMR spectroscopy / G. M. Clore [et al.] // Biochemistry. — 1990. — Vol. 29. — P. 87—91.

280. Carbon-13 chemical shift anisotropies of solid amino acids / C. Ye [et al.] // Magnetic Resonance in Chemistry. — 1993. — Aug. — Vol. 31, no. 8. — P. 699—704.

281. Characterization of fibril dynamics on three timescales by solid-state NMR / A. A. Smith [et al.] // J. Biomol. NMR. — 2016. — Vol. 65, no. 3. — P. 171—191.

282. Correction to: characterization of fibril dynamics on three timescales by solid-state NMR / A. A. Smith [et al.] // J. Biomol. NMR. — 2018. — Vol. 70, no. 3. — P. 203—203.

283. Slow motions in microcrystalline proteins as observed by MAS-dependent ^{15}N rotating-frame NMR relaxation / A. Krushelnitsky [et al.] // J. Magn. Reson. — 2014. — Nov. — Vol. 248. — P. 8—12.

284. Microsecond motions probed by Near-Rotary-Resonance $R_{1\rho}$ ^{15}N MAS NMR experiments : the model case of protein overall-rocking in crystals / A. Krushelnitsky [et al.] // J. Biomol. NMR. — 2018. — Vol. 71, no. 1. — P. 53—67.

285. The structure of human ubiquitin in 2-methyl-2, 4-pentanediol: A new conformational switch / K. Y. Huang [et al.] // Prot. Sci. — 2011. — Vol. 20, no. 3. — P. 630—639.

286. Berendsen, H. The missing term in effective pair potentials / H. Berendsen, J. Grigera, T. Straatsma // J. Phys. Chem. — 1987. — Vol. 91, no. 24. — P. 6269—6271.

287. Simulations of a protein crystal with a high resolution X-ray structure: evaluation of force fields and water models / D. S. Cerutti [et al.] // Journal of Physical Chemistry. — 2010. — Vol. 114, no. 40. — P. 12811—12824.

288. AMBER 16, University of California, San Francisco / D. Case [et al.]. — University of California, San Francisco, 2015.

289. AMBER 11, University of California, San Francisco / D. Case [et al.]. — University of California, San Francisco, 2010.

290. Comparison of multiple Amber force fields and development of improved protein backbone parameters / V. Hornak [et al.] // Proteins. — 2006. — Vol. 65, no. 3. — P. 712—725.

291. Improved side-chain torsion potentials for the Amber ff99SB protein force field / K. Lindorff-Larsen [et al.] // Proteins. — 2010. — Vol. 78, no. 8. — P. 1950—1958.

292. Crystallographic analysis of metal-ion binding to human Ubiquitin / F. Arnesano [et al.] // Chemistry European Journal. — 2011. — Vol. 17, no. 5. — P. 1569—1578.

293. Structural probing of Zn(ii), Cd(ii) and Hg(ii) binding to human ubiquitin / G. Falini [et al.] // Chemical Communications. — 2008. — Vol. 45. — P. 5960—5962.

294. Dreydoppel, M. Slow ring flips in aromatic cluster of GB1 studied by aromatic ^{13}C relaxation dispersion methods / M. Dreydoppel, H. N. Raum, U. Weininger // J. Biomol. NMR. — 2020. — Vol. 74, no. 2. — P. 183—191.

295. The penultimate rotamer library / S. C. Lovell [et al.] // Proteins. — 2000. — Vol. 40, no. 3. — P. 389—408.

296. Rovó, P. Recent advances in solid-state relaxation dispersion techniques / P. Rovó // Solid State Nucl. Magn. Reson. — 2020. — T. 108. — C. 101665.

297. Zhang, R. Proton chemical shift tensors determined by 3D ultrafast MAS double-quantum NMR spectroscopy / R. Zhang, K. H. Mroue, A. Ramamoorthy // J. Chem. Phys. — 2015. — Vol. 143, no. 144201. — P. 1—6.

298. Chen, J. Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation / J. Chen, C. L. Brooks, P. E. Wright // J. Biomol. NMR. — 2004. — Vol. 29, no. 3. — P. 243—257.

299. Halle, B. Model-free analysis of stretched relaxation dispersions / B. Halle, H. Jóhannesson // J. Magn. Reson. — 1998. — Vol. 135, no. 1. — P. 1—13.

300. Deviations from the simple two-parameter model-free approach to the interpretation of ^{15}N nuclear magnetic relaxation of proteins / M. G. Clore [et al.] // Journal of the American Chemical Society. — 1990. — Vol. 112, no. 12. — P. 4989—4991.

301. Skrynnikov, N. R. Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond timescale side-chain motions / N. R. Skrynnikov, O. Millet, L. E. Kay // Journal of the American Chemical Society. — 2002. — Vol. 124, no. 22. — P. 6449—6460.

302. Kowalewski, J. Dipole–dipole coupling constant for a directly bonded CH pair-A carbon-13 relaxation study / J. Kowalewski, M. Effemey, J. Jokisaari // J. Magn. Reson. — 2002. — Vol. 157, no. 2. — P. 171—177.

303. NMR Determination of amide N–H equilibrium bond length from concerted dipolar coupling measurements / L. Yao [et al.] // Journal of the American Chemical Society. — 2008. — Vol. 130, no. 49. — P. 518—520.

304. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors / L. Cao [et al.] // Science. — 2020. — Vol. 370, no. 6515. — P. 426—431.

305. Lebedenko, O. O. Order/Disorder Transitions Upon Protein Binding: A Unifying Perspective / O. O. Lebedenko, A. Sekhar, N. R. Skrynnikov // Proteins: Structure, Function, and Bioinformatics. — 2024. — Aug.

306. Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory / D. E. Koshland // Angewandte Chemie International Edition in English. — 1995. — Jan. — Vol. 33, no. 23/24. — P. 2375—2378.

307. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations / A. Raval [et al.] // Journal of proteins structure, function and bioinformatics. — 2012. — May. — Vol. 80, no. 8. — P. 2071—2079.

308. Nerenberg, P. S. New developments in force fields for biomolecular simulations / P. S. Nerenberg, T. Head-Gordon // Current Opinion in Structural Biology. — 2018. — Апр. — Т. 49. — С. 129—138.

309. Zhu, J. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states / J. Zhu, X. Salvatella, P. Robustelli // Nature Communications. — 2022. — Oct. — Vol. 13, no. 1.

310. Diffusing protein binders to intrinsically disordered proteins / C. Liu [et al.] // bioRxiv. — 2024. — July.