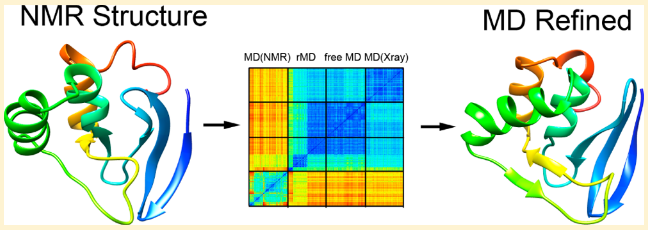
Protocol To Make Protein NMR Structures Amenable to Stable Long Time Scale Molecular Dynamics Simulations

Da-Wei Li and Rafael Brüschweiler*

Campus Chemical Instrument Center and Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306, United States

ABSTRACT: A robust protocol for the treatment of NMR protein structures is presented that makes them amenable to long time scale molecular dynamics (MD) simulations that are stable. The protocol embeds an NMR structure in a native low energy region of the recently developed ff99SB_$\varphi\psi$(g24;CS) molecular mechanics force field. Extended MD trajectories that start from these structures show good consistency with proton-proton nuclear Overhauser effect data, and they reproduce NMR chemical shift data better than the original NMR structures as is demonstrated for four protein systems. Moreover, for all proteins studied here the simulations spontaneously approach the X-ray crystal structures, thereby improving the effective resolution of the initial structural models.



INTRODUCTION

Molecular dynamics (MD) computer simulations of proteins provide important new insights into protein behavior and function.[1] In recent years, all-atom molecular dynamics (MD) simulations of proteins in explicit solvent have made major strides in their ability to quantitatively reproduce a wide range of experimental structural dynamic parameters of proteins. These advances were made possible by the development of improved all-atom molecular mechanics force fields and the availability of ever more powerful computer hardware, which have turned microsecond time scale simulations in explicit solvent practically feasible.[2] The introduction of the AMBER ff99SB,[3] AMBER ff03,[4] and CHARMM CMAP[5,6] force fields represented important steps in force-field development, which triggered the use of different types of experimental protein data for the quantitative certification of MD simulations and their further improvement. In particular, the validation of free MD trajectories of proteins against solution NMR measurements of intact proteins and peptides, such as residual dipolar couplings,[7−9] scalar *J*-couplings,[8,10,11] spin relaxation order parameters,[12−14] and chemical shifts,[15,16] demonstrate that the simulations can reach semiquantitative or even quantitative agreement with experiment rendering these simulations remarkably realistic. Recently, we optimized the ff99SB backbone dihedral angle potentials using NMR chemical shift data of intact full-length proteins. Combined with the ILDN side-chain parameters,[17] it resulted in MD trajectories that are in very good agreement with experimental data for a diverse set of proteins.[18]

A requirement for the success of MD, in addition to an accurate force field and powerful computers, is the availability of a suitable three-dimensional protein structure as a starting point. The vast majority of MD starting structures have been determined by X-ray crystallography with a resolution preferably ∼2 Å or better. Although over 8500 protein structures determined by NMR spectroscopy have been deposited in the Protein Data Bank (PDB),[19] because the quality of NMR structures is on average lower and more heterogeneous,[20] NMR models are often not the first choice for protein structure data mining[21] and MD simulations.[22] In our own experience, they have a tendency to lead to unstable MD trajectories that drift rapidly away from the initial structure. This situation is unfortunate since for many of these proteins no crystal structure is available. Although the initial NMR structures are typically consistent with the majority of the NMR constraints used during their generation, such as nuclear Overhauser effect (NOE) derived distance constraints and residual dipolar couplings (RDCs), this does not guarantee that they occupy an energy minimum of a current force field. This is particularly relevant if the NMR structures were determined without the use of a realistic force field, such as using a basic repulsive function for the van der Waals terms and without Coulomb interactions, leading to ensembles with distorted covalent geometries. In the presence of a physical force field, such structures experience forces that lead to unstable protein behavior during MD simulations drifting away from the initial structure, which can affect major portions of the protein. To prevent such unstable behavior and to make NMR structures amenable to standard MD simulations, a protocol is presented here that suitably embeds NMR structures in the force field permitting long time scale MD simulations into the submicro-

Received: December 9, 2013
Published: February 14, 2014

second range without adversely affecting the quality of the structure. In fact, for the systems studied here, the quality of the structure improves over the initial structure as judged from NMR data not used during structure refinement. Therefore, this work also provides an improved refinement protocol for protein structures determined by NMR.

■ METHODS

Protein Selection. All proteins selected for this study, which are listed in Table 1, fulfill the following two criteria: (1) a NMR structure is available along with the experimental NOE constraint list, chemical shifts, and residual dipolar couplings; (2) an X-ray crystal structure is available at a resolution of 2 Å or better. The protein PDB codes,[19] amino acid sequence lengths, Biological Magnetic Resonance Data Bank (BMRB) chemical shift access codes,[23] and mrblock codes from the NMR Restraint Grid[24] are given in Table 1. For protein 2KZN, the N-terminus (residues Met1 to Gln39) were excluded because NOE data are very sparse for this region.

Table 1. List of Proteins with Their Database Entry Codes

protein name	NMR PDB code	X-ray PDB code	amino acid sequence length	BMRB entry	mrblock
hen egg lysozyme	1E8L	1DPX	129	4831	516303
peptide methionine sulfoxide reductase msrB	2KZN	3E0O	141	5619	472085
PDZ2 domain of human phosphatase	3PDZ	3LNX	94	4123	454978[a]
transcription factor Mbp1	1L3G	1MB1	98	4254	385599

[a] RDC data were obtained from Prof. Andrew Lee.

MD Simulations. All MD simulations were performed using the Gromacs 4.5 package[25] with the ff99SB_φψ-(g24;CS)[26] + ILDN[17] force field. Water molecules were explicitly represented using the TIP3P model.[27] The integration time step was set to 2 fs with all bond lengths involving hydrogen atoms constrained by the SETTLE algorithm. Electrostatic interactions were cut off at 10 Å and the long-range electrostatic interactions were calculated using the PME algorithm with 1.2 Å spacing; van der Waals interactions were cut off at 8 Å. Standard minimization procedures described previously[28] were applied first before each system was simulated at 300 K for 1 ns with all protein heavy atoms positionally restrained by a harmonic potential, followed by another 1 ns simulation with the positional restraints removed. The final production runs were simulated at constant pressure (1 atm) and constant temperature (at 300, 330, and 360 K) for a duration of 100−200 ns. The total trajectory lengths are dictated by the protein behavior with longer trajectories favored for proteins that display unstable or drifting behavior over the first 100 ns. For the MD simulations using NMR structures as starting points, the first NMR structure deposited in the PDB was used.

NOE Restraint Treatment. NOE distance restraints were included in the MD simulation as an energy term defined as follows:

$$V = \begin{cases} 0 & \text{if } r < r_1 \\ 0.5k(r - r_1)^2 & \text{if } r_1 \le r < r_2 \\ 0.5k(r_2 - r_1)(2r - r_2 - r_1) & \text{if } r \ge r_2 \end{cases} \tag{1}$$

where V is the potential energy, r is the distance between the pair of atoms, r_1 is the NOE-derived upper limit, $r_2 = r_1 + 2$ Å, and k is the force constant. For a NOE between a proton X and N symmetry-related protons $n = 1, 2, ..., N$, such as the three protons of a methyl group, with distances r_n, an apparent distance r was used for eq 1 given by

$$r = [\sum_{n=1}^{N} r_n^{-6}]^{-1/6} \tag{2}$$

In the case of ambiguous assignments, e.g., for methylene protons, all potential proton pairs were required to fulfill the distance constraints whereby the upper limit r_1 was increased by the (known) interatomic distances between the ambiguous protons. The same protocol was applied for the back-calculation of NOEs for ensemble validation. All NOE restraints were taken from the FRED database of the NMR Restraints Grid, which contains sets of NMR constraints that have been converted to a standard format and filtered to eliminate misassigned NOEs.[29]

Chemical Shift Analysis. The quality of conformational ensembles was assessed by comparing chemical shifts back-calculated from ensembles with their experimental counterparts. For this purpose, chemical shifts were back-calculated either from individual MD snapshots or from each structure from the NMR ensemble using both the PPM[30] and Shifts[31,32] software followed by averaging over all members of the ensemble with equal weight. Because PPM was specifically parametrized for the prediction of chemical shifts from large MD ensembles, i.e., not from single structures, the program Shifts was used to assess the deposited NMR PDB structures. The chemical shift difference between experimental values, taken from the BMRB,[23] and the ensemble averaged predicted chemical shifts was expressed as a chemical shift root-mean-square deviation (rmsd in units of parts per million) as described previously.[15]

RDC Analysis. For each alignment medium, an alignment tensor was fitted to the MD or NMR ensembles by singular value decomposition (SVD) after all structures were aligned with respect to the first snapshot or first structure in the case of NMR ensembles.[7] Residual dipolar couplings were then back-calculated and compared with their experimental counterparts via Q values as described previously.[18,33] Experimental RDC data were obtained from the BMRB Restraints Grid for all proteins, except for the PDZ domain 3PDZ (data provided by Dr. Andrew Lee).

■ RESULTS

Three of the four NMR structures, namely 1E8L, 3PDZ, and 1L3G, had previously been solved using traditional NMR methods using large data sets of NOEs and scalar[3] *J*-couplings as energy terms. Consequently, the deposited NMR structures had very few NOE violations. For two of the proteins (1E8L and 1L3G), in addition, backbone N−H RDCs were used as orientational constraints during structure determination, which led to structures with very low RDC Q values of 0.08 and 0.03, respectively. Cross-validation performed with NMR data that were excluded during structure determination, namely the

Table 2. Quantitative Assessment of Different Protein Ensembles Based on the Comparison of Experimental and Back-Calculated NMR Data

protein system	assessment	NMR PDB	NMR MD	free MD	X-ray MD
1E8L, 1DPX	RDC Q	0.08	0.35	0.26	0.26
	chem shift rmsd[a]				
	Cα	2.10 (1.76)	1.09 (1.49)	1.05 (1.43)	1.04 (1.37)
	Cβ	2.69 (1.95)	1.51 (1.73)	1.55 (1.61)	1.38 (1.60)
	C′	2.44 (2.09)	1.29 (1.73)	1.24 (1.80)	1.22 (1.77)
	NOE violations[b]	0 (0)	66 (38)	59 (36)	45 (29)
2KZN, 3E0O	RDC Q	0.17	0.63	0.37	0.42
	chem shift rmsd[a]				
	Cα	1.59 (1.47)	1.84 (1.92)	1.67 (1.71)	1.30 (1.41)
	Cβ	1.70 (1.73)	1.99 (1.95)	1.75 (1.82)	1.60 (1.61)
	C′	1.63 (1.93)	1.53 (1.87)	1.33 (1.71)	1.33 (1.66)
	NOE violations[b]	68 (50)	343 (155)	280 (136)	314 (131)
3PDZ, 3LNX	RDC Q	0.84	0.58	0.33	0.29
	chem shift rmsd[a]				
	Cα	2.44 (2.13)	2.24 (2.16)	2.01 (1.98)	1.92 (1.84)
	Cβ	2.29 (2.43)	2.26 (2.40)	2.29 (2.34)	2.05 (2.23)
	NOE violations[b]	25 (20)	641 (206)	331 (153)	314 (149)
1L3G, 1MB1	RDC Q	0.03	0.65	0.41	0.37
	chem shift rmsd[a]				
	Cα	2.16 (1.67)	1.58 (1.68)	1.22 (1.43)	1.04 (1.19)
	Cβ	2.59 (1.77)	1.29 (1.52)	1.11 (1.43)	1.07 (1.37)
	NOE violations[b]	0 (0)	99 (43)	84 (33)	83 (32)

[a] Chemical shifts are predicted using PPM and Shifts (the Shifts results are in parentheses). [b] Values are the sum of all NOE violations >1 Å. A NOE violation is defined as the r^{-6}-weighted average of the distance calculated from the ensemble minus the corresponding experimental upper limit. Numbers in parentheses are the total number of violations that are larger than 1 Å.

chemical shifts for all three proteins in addition to RDCs for protein 3PDZ, show rather large deviations between back-calculated and experimental values. This suggests that different types of NMR data, such as NOEs and chemical shifts, are either mutually inconsistent or that these structures are underdetermined with respect to the NMR data used during structure calculation. Based on the results described below, the first possibility can be excluded. The remaining NMR structure 2KZN had been solved using CS-ROSETTA[34] using sparse NOE restraints obtained from deuterated ILV-methyl protonated samples, RDCs, and chemical shifts. The resulting NMR ensemble had few NOE violations, low RDC Q values, and low chemical shift rmsd.

We ran 100 ns of unrestrained MD simulations in explicit solvent using the first model of each NMR PDB file as the starting structure with the resulting trajectories termed “NMR MD”. The protein structures moved quickly away from the original structures, and the back-calculated NMR data from the trajectories were systematically different from those calculated from the initial structures. This behavior indicates that the deposited NMR structures do not occupy regions of minimal energy of the molecular mechanics force field. Back-calculated NMR data were then compared with experiment (Table 2). The MD trajectories had relatively large RDC Q values, large numbers of NOE violations, and large chemical shift errors. As a control, we also ran 100 ns MD simulations using the X-ray crystal structures as initial structures, termed “X-ray MD”, and assessed the resulting ensembles in terms of their NMR properties. The discrepancies between “X-ray MD” and experiment were substantially smaller than those between “NMR MD” and experiment (Table 2, last column), which suggests that the X-ray structures of the proteins studied here represent better MD starting structures than the NMR structures.

The flowchart of our refinement protocol is depicted in Figure 1. To refine the NMR structures by restrained MD, we first ran MD simulations for six different combinations of simulation temperatures (300, 330, and 360 K) and NOE restraint force constants k (see eq 1) (30 and 100 kJ/mol/nm^2) for 100 or 200 ns length for each protein, which are termed

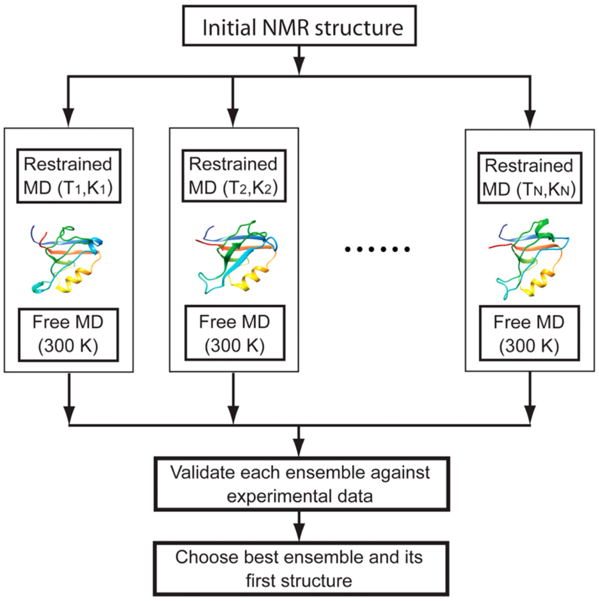


Figure 1. Flowchart of computational protocol for the refinement of original NMR structures to make them amenable to extended restraint-free MD simulations.

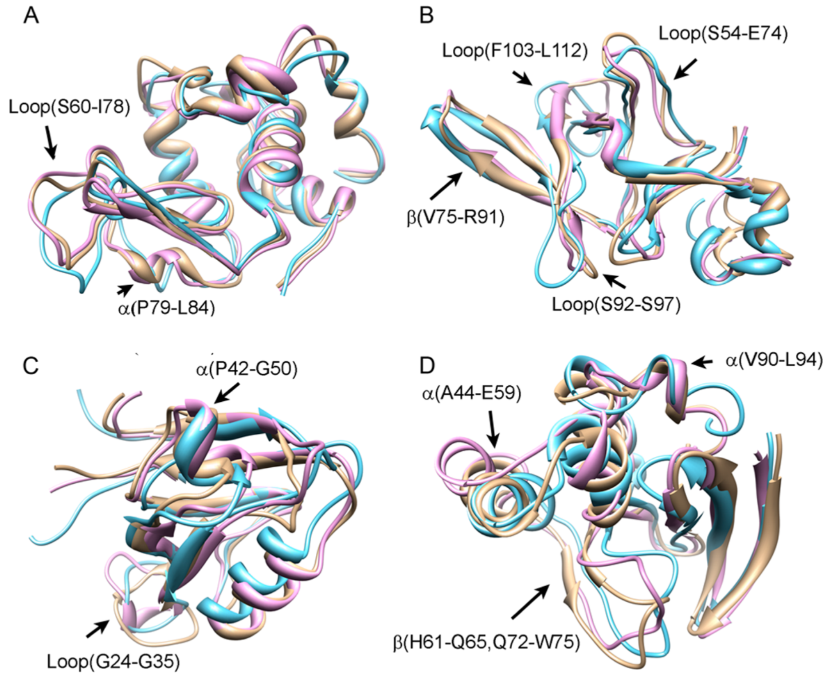


Figure 2. Ribbon diagrams of proteins (A) 1E8L, (B) 2KZN, (C) 3PDZ, and (D) 1L3G. The X-ray crystal structures, the refined NMR structures, and the initial NMR structures are depicted in gold, magenta, and cyan, respectively. The refined NMR structures (magenta) are on average much closer to the X-ray crystal structures (gold) than to the initial NMR structures (cyan).

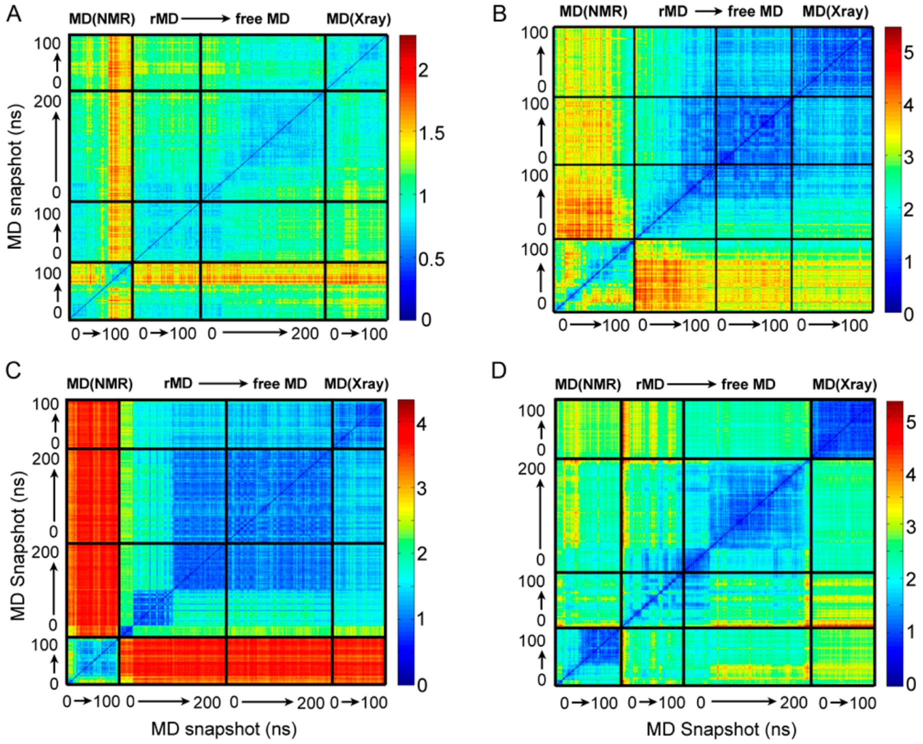


Figure 3. Pairwise backbone rmsd's of all snapshots of the NMR MD ensemble "MD(NMR)", the restrained MD ensemble "rMD", the free MD ensemble "free MD", and the X-ray MD ensemble "MD(Xray)" of proteins (A) 1E8L, (B) 2KZN, (C) 3PDZ, and (D) 1L3G. The lengths of the MD trajectories are indicated along the axes of the figures. The pairwise backbone rmsd's are color coded using the individual color maps depicted on the right-hand side of each panel.

"restrained MD". Next, the final conformation of each of the six restrained MD trajectories corresponds to a refined NMR structure candidate, which is taken as a starting structure for a 100 or 200 ns long free MD simulation (termed "free MD") in the *absence* of any experimental restraints. This protocol was repeated between two and five times per protein with different randomly assigned initial atomic velocities. Chemical shifts, NOE constraints, and RDCs were back-calculated from the free MD trajectories and compared with experiment to determine the highest quality MD ensemble. The higher number of iterations is warranted when the agreement is poor. The refined NMR structure candidate that served as the starting point of the highest quality free MD ensemble is then designated as the "refined NMR structure". Without exception, the quality of the refined NMR structures improved substantially compared with the original NMR model, as evidenced by the fact that the free MD ensembles obtained from the refined structures always had lower RDC Q values, lower chemical shift errors, and fewer NOE violations than the NMR MD ensembles (Table 2). The quality of the refined structures is comparable to or only slightly worse than the corresponding X-ray structures, as shown by the errors between back-calculated and experimental NMR data.

In Figure 2, the first model of the NMR structure, the X-ray crystal structure, and the refined NMR structure are superimposed for all four proteins. Figure 2 shows that the refined NMR structures (magenta) resemble the X-ray structures (gold) much more than the original NMR structures (cyan). For protein 1E8L (Figure 2A), the structure of loop S60−I78 undergoes substantial changes. Immediately following this loop, the six residues P79−L84 form a short α-helix in the refined structures. Both of these structural elements become more similar to the crystal structure. For protein 2KZN (Figure 2B), conformations and inter-residue contact patterns significantly change in the three loop regions S92−S97, S54−E74, and F103−L112. The lengths of both β-strands of the β-hairpin (V75−R91) shorten in the refined structure. For all these regions, the refined structures become more similar to the crystal structure. For protein 3PDZ (Figure 2C), the main change of conformation occurs for the short α-helix P42−G50. This helix is unstable during the MD simulations that start from the NMR structure, but it is stable during the MD simulations for both the crystal and refined NMR structures. The bottom-left corner loop (G24−G35) also displays large differences between the three structural models. This part is flexible in all three MD simulations, but does not converge toward a common distribution. Protein 1L3G (Figure 2D) represents a somewhat more challenging case. The C-terminus of the refined structure (V90−L94) becomes more similar to the crystal structure. However, in one of the regions labeled in Figure 2D, both the initial NMR structure and the refined NMR structure fail to form the β-sheet found in the crystal structure (β-strands H61−Q65 and Q72−W75) and, as a consequence, the loop conformation (Q65−Q72) also differs. The packing of the α-helix (A44−E59) of the NMR structure and the refined structure also differ with respect to the crystal structure.

To quantitatively compare the ensembles generated by different MD runs, we computed the pairwise rmsd's of all backbone heavy atoms among all four ensembles with the result visualized in Figure 3. Such plots provide direct insight into the presence and stability of free energy basins sampled during a trajectory. Due the low pairwise rmsd's of conformations visited over a certain MD segment, such basins emerge in Figure 3 as blue squares along the diagonal. On the other hand, if the protein visits substantially different regions of conformation space with large pairwise rmsd's, yellow or red off-diagonal rectangles appear. Figure 3 shows that the "NMR MD" ensembles (lower left part of each plot) are not particularly stable and significantly different from the "restrained MD", "free MD", and "X-ray MD" ensembles. On the other hand, the "free MD" ensembles are remarkably similar to the "X-ray MD" ensembles for Figure 3A−C, although no information about the X-ray structures entered the "free MD" simulations. This resemblance is a consequence of the combination of the physical accuracy of the MD force field and the application of relatively weak NOE-distance constraints during restrained MD. For protein 1L3G (Figure 3D), the "free MD" ensemble does not get closer than ~2 Å backbone rmsd to the "X-ray MD" ensemble. The difference between the two ensembles is likely to be caused by the absence of NOEs in a large protein segment (A44−W75). The final backbone heavy atom rmsd between the free MD ensembles and the X-ray MD ensembles for the four proteins are 1.1 (1E8L), 1.5 (2KZN), 1.9 (3PDZ), and 2.3 Å (1L3G) with an average of 1.7 Å. Hence, 1L3G is the only protein whose refined structure has a backbone rmsd with respect to the X-ray structure that exceeds 2 Å. Still, based on the back-calculated NMR data, the quality of the final structure of 1L3G is clearly better than that of the initial NMR model.

■ DISCUSSION

The availability of high-resolution protein structures is crucial for understanding the intramolecular forces that determine their stability and their interactions with other proteins, nucleic acids, substrates, and drugs. For the vast majority of protein structures in the PDB determined by X-ray crystallography or by solution NMR spectroscopy, the accuracy of structural models can be limited: in the case of crystal structures because of packing effects and static disorder and in the case of NMR structures due to limited sensitivity leading to overconstrained or underconstrained protein regions. In traditional NMR structure determination, a large number of NMR experimental constraints is required, such as NOE-derived distance constraints, dihedral angle constraints from scalar J-couplings, and orientational constraints from RDCs. Using integrated software packages, such as CYANA,[55] XPLOR-NIH,[56] and ARIA,[35] these constraints are then converted to pseudoenergy terms, which are added to the physical force field containing energy terms that reflect bond lengths, bond angles, and van der Waals interactions. The programs then generate structures that are consistent with an optimal number of constraints by minimizing the total energy. During this process, the energy terms converted from the constraints typically have a large weight so that they dominate the energies of the physical force field.

While for X-ray crystal structures the resolution of the diffraction data and the R-factor R_{free} provide direct measures for their accuracy, the quality of NMR structures depends in a more complex way on the amount and quality of the data.[36] This makes the assessment of their accuracy harder. It has been estimated that high-quality NMR structures have accuracies comparable to 2.0 Å resolution X-ray crystal structures, whereas average structures have resolutions that can be considerably worse.[20,37,38] Only a few NMR structures have been determined at even higher resolution since the amount and quality of experimental data required exceeds what can be routinely collected.

Molecular dynamics has the potential to improve the accuracy of protein structures that have been determined by experiment or predicted by bioinformatics approaches.[39,40] The use of free MD simulations combined with accelerated sampling technologies using physics-based force fields have shown promise in protein structure refinement[41−43] and prediction.[44,45] Recently, we found that the ff99SB_$\varphi\psi$(g24;CS) force field is sufficiently accurate to permit direct refinement of low-resolution protein structures via free MD simulation on the microsecond time scale, provided that sampling is not restricted by a high free energy barrier to the native conformation.[46] Although MD simulations into the microsecond time scale are becoming routine for many proteins, unconstrained brute force MD simulations are often still too short for direct structure refinement, even at moderately elevated temperatures.[46] In order to speed up sampling without distorting the native target structure itself, we use restrained MD simulations at multiple temperatures from 300 to 360 K and different force constants for the NOE-distance restraints. In this way, experimental NMR structures are gently embedded in the underlying physical force field preventing rapid large-scale excursions from the initial structures. The simulations are then continued over 100−200 ns to obtain protein structures that are further relaxed, populating the low free energy region of the physical force field. Experimental structures treated in this way are then amenable to long-term MD simulations to explore in silico their biophysical properties at ambient conditions. Our method has features in common with previous NMR structure refinement protocols. However, previous MD refinement simulations were often very short and used notably high constraint force constants.[22,47,48] By contrast, we put more emphasis on the accuracy of the physical force field and run much longer MD simulations to allow full relaxation of the structural models. The final restraint-free MD cycle and the selection criteria ensure that our models are consistent with both the physical force field and the NMR constraints.

The MD simulations in the presence of NOE restraints represent, like all MD simulations, (pseudo)stochastic processes. Hence, trajectories starting from identical initial conformations, but with different initial velocities, can reach different free energy basins over the 100−200 ns trajectory lengths. In order to provide each protein the chance to sample a variety of possibilities to become embedded in the underlying force field, the protocol uses multiple restrained MD simulations at different temperatures (300−360 K) and force constants for the restraints. In fact, the four proteins studied here display different behaviors during refinement and different requirements for optimal refinement. For protein 1E8L the restrained MD simulations at any of the temperatures (300, 330, or 360 K) displayed a high probability to reach a nativelike state within 100 ns. By contrast, for protein 3PDZ the simulation temperature proved critical for successful refinement. The MD simulations at 300 or 330 K failed to reach an acceptable structure, but when the temperature was raised to 360 K, most of the refinement MD simulations reached a nativelike state within 100 ns. Proteins 2KZN and 1L3G both reflect a low propensity for refinement with a success rate of as low as ~20% in the case of protein 2KZN, making a large number of independent simulations important for successful refinement. For protein 1L3G, the best structure refined still had a rmsd ~ 2 Å with respect to the X-ray crystal structure. Such diverse protein behavior reflects energy barriers between the initial structures and the native conformations that vary for different proteins and their NMR structures. Another factor is the availability of a complete set of long-range NOEs. For example, for protein 2KZN only a sparse set of NOE restraints is available, which might explain why 2KZN proved to be a difficult target for refinement by standard restrained MD simulations. For protein 1L3G, many NOEs in the region between residues Lys45 and Gly73 were missing, which is the likely reason why the refined structure is not as close to the X-ray structure as for the other proteins. In fact, Figure 2D shows that the main difference between the refined NMR and the X-ray structure is in the region between Ala44 and Trp75, which contains the α-helix and the loop in the left part of Figure 2D. The diverse protein behavior observed in restrained MD simulations implies that a practically useful approach in protein structure refinement is to employ several MD simulations under various conditions, i.e., different temperatures and restraint force constants, followed by selection of the best structure determined on the basis of back-calculated NMR data that were not used during the structure determination process. In practice, we suggest that the protocol is repeated initially twice; additional simulations are added if the refined results are not optimal as judged from a stability analysis of the trajectories and back-calculated NMR data.

Our strategy for the treatment of NMR structures differs in scope from other strategies described in the literature for the representation of proteins in terms of conformational ensembles. These include time-averaged restrained MD,[49] replica exchange simulations in the presence of NMR constraints,[50] and conformational ensemble reconstruction.[51−53] In these approaches, the primary goal is to either directly use or validate experimental constraints to generate an ensemble that is consistent with the experimental data. By contrast, the goal of the present work is to embed NMR structures in a physically accurate force field so that it can be studied by free MD simulations in full analogy to MD studies of X-ray crystal structures.

An increasing number of proteins is now available in the PDB that have structures that were solved both by NMR and by X-ray crystallography. A recent survey of such proteins found that, for many proteins, NMR and X-ray structures differ from each other significantly.[54] The different environments, especially crystal packing effects, are often considered to be the main reason behind the structural difference. For the proteins studied here, we find that at least some of the structural differences might stem from systematic biases of the NMR model. This suggests that, unless systematic differences between crystal and solution structures are validated by independent measurements, such as RDC or chemical shifts, observed differences between NMR and crystal structures need to be interpreted with caution. Protein structures that are generated by the MD-based NMR structure refinement protocol described here should always be assessed through comparison between experimental and back-calculated NMR data to guarantee that these refined structures are of comparable or better quality when compared with the initial NMR protein structures.

With the computational power available today, the application of our protocol takes only a fraction of the total time for protein sample preparation and NMR data collection. This should make the addition of the (sub-)microsecond MD refinement step to standard NMR structure determination protocols affordable and practically useful.

■ CONCLUSION

The quality of MD simulations is now reaching a level which permits protein structure refinement at nearly experimental accuracy. Previously, we found that due to the presence of sampling-related bottlenecks this type of structure refinement only works for a subset of systems. In this work, we provide a protocol that addresses this challenge. With the proper choice of simulation temperature and NOE restraint strength, the MD simulations make the structures nativelike within less than a microsecond simulation time, thereby considerably improving the quality of the structure.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bruschweiler.1@osu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Andrew Lee for providing the RDC data of the PDZ domain. This work was supported by the National Science Foundation (Grant MCB-1360966).

■ REFERENCES

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.

(2) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120.

(3) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.

(4) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.

(5) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, *90*, L36.

(6) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257.

(7) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158.

(8) Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys. J.* **2010**, *99*, 647.

(9) Long, D.; Li, D. W.; Walter, K. F. A.; Griesinger, C.; Brüschweiler, R. *Biophys. J.* **2011**, *101*, 910.

(10) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009**, *97*, 853.

(11) Markwick, P. R.; Showalter, S. A.; Bouvignies, G.; Brüschweiler, R.; Blackledge, M. *J. Biomol. NMR* **2009**, *45*, 17.

(12) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.

(13) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724.

(14) Trbovic, N.; Kim, B.; Friesner, R. A.; Palmer, A. G. *Proteins* **2008**, *71*, 684.

(15) Li, D. W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246.

(16) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 1220.

(17) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950.

(18) Li, D. W.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778.

(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.

(20) Bagaria, A.; Jaravine, V.; Guntert, P. *Comput. Biol. Chem.* **2013**, *46*, 8.

(21) Sheffler, W.; Baker, D. *Protein Sci.* **2009**, *18*, 229.

(22) Nederveen, A. J.; Doreleijers, J. F.; Vranken, W.; Miller, Z.; Spronk, C. A. E. M.; Nabuurs, S. B.; Guntert, P.; Livny, M.; Markley, J. L.; Nilges, M.; Ulrich, E. L.; Kaptein, R.; Bonvin, A. M. J. J. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 662.

(23) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402.

(24) Doreleijers, J. F.; Vranken, W. F.; Schulte, C.; Lin, J. D.; Wedell, J. R.; Penkett, C. J.; Vuister, G. W.; Vriend, G.; Markley, J. L.; Ulrich, E. L. *J. Biomol. NMR* **2009**, *45*, 389.

(25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(26) Li, D. W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773.

(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(28) Li, D. W.; Meng, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 14610.

(29) Doreleijers, J. F.; Nederveen, A. J.; Vranken, W.; Lin, J. D.; Bonvin, A. M. J. J.; Kaptein, R.; Markley, J. L.; Ulrich, E. L. *J. Biomol. NMR* **2005**, *32*, 1.

(30) Li, D. W.; Brüschweiler, R. *J. Biomol. NMR* **2012**, *54*, 257.

(31) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321.

(32) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408.

(33) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563.

(34) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685.

(35) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. *Bioinformatics* **2007**, *23*, 381.

(36) Rosato, A.; Tejero, R.; Montelione, G. T. *Curr. Opin. Struct. Biol.* **2013**, *23*, 715.

(37) Billeter, M. *Q. Rev. Biophys.* **1992**, *25*, 325.

(38) Montelione, G. T.; Zheng, D. Y.; Huang, Y. P. J.; Gunsalus, K. C.; Szyperski, T. *Nat. Struct. Biol.* **2000**, *7*, 982.

(39) Nilges, M.; Brunger, A. T. *Protein Eng.* **1991**, *4*, 649.

(40) Vieth, M.; Kolinski, A.; Brooks, C. L.; Skolnick, J. *J. Mol. Biol.* **1994**, *237*, 361.

(41) Fan, H.; Mark, A. E. *Protein Sci.* **2004**, *13*, 211.

(42) Chen, J. H.; Brooks, C. L. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 922.

(43) Ishitani, R.; Terada, T.; Shimizu, K. *Mol. Simul.* **2008**, *34*, 327.

(44) Shell, M. S.; Ozkan, S. B.; Voelz, V.; Wu, G. H. A.; Dill, K. A. *Biophys. J.* **2009**, *96*, 917.

(45) Voelz, V. A.; Shell, M. S.; Dill, K. A. *PLoS Comput. Biol.* **2009**, *5*, No. e1000281.

(46) Li, D. W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2012**, *8*, 2531.

(47) Bertini, I.; Case, D. A.; Ferella, L.; Giachetti, A.; Rosato, A. *Bioinformatics* **2011**, *27*, 2384.

(48) Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317.

(49) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289.

(50) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, 138.

(51) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3.

(52) Clore, G. M.; Schwieters, C. D. *Biochemistry* **2004**, *43*, 10678.

(53) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.

(54) Andrec, M.; Snyder, D. A.; Zhou, Z. Y.; Young, J.; Montelione, G. T.; Levy, R. M. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 449.

(55) López-Méndez, B.; Güntert, P. *J. Am. Chem. Soc.* **2006**, *128*, 13112−13122.

(56) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Res.* **2003**, *160*, 66−74.

■ NOTE ADDED AFTER ASAP PUBLICATION

Due to a production error, this paper was published ASAP on February 26, 2014 with two missing references. The revised version was reposted on March 5, 2014.