JCTC Journal of Chemical Theory and Computation

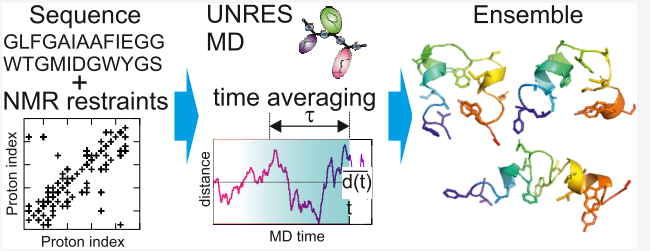
pubs.acs.org/JCTC Article

Implementation of Time-Averaged Restraints with UNRES Coarse-Grained Model of Polypeptide Chains

Nguyen Truong Co, Cezary Czaplewski, Emilia A. Lubecka, and Adam Liwo*

Cite This: J. Chem. Theory Comput. 2025, 21, 1476−1493 Read Online

ABSTRACT: Time-averaged restraints from nuclear magnetic resonance (NMR) measurements have been implemented in the UNRES coarse-grained model of polypeptide chains in order to develop a tool for data-assisted modeling of the conformational ensembles of multistate proteins, intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs), many of which are essential in cell biology. A numerically stable variant of molecular dynamics with time-averaged restraints has been introduced, in which the total energy is conserved in sections of a trajectory in microcanonical runs, the bath temperature is maintained in canonical runs, and the time-average-restraint-force components are scaled up with the length of the memory window so that the restraints affect the simulated structures. The new approach restores the conformational ensembles used to generate ensemble-averaged distances, as demonstrated with synthetic restraints. The approach results in a better fitting of the ensemble-averaged interproton distances to those determined experimentally for multistate proteins and proteins with intrinsically disordered regions, which puts it at an advantage over all-atom approaches with regard to the determination of the conformational ensembles of proteins with diffuse structures, owing to a faster and more robust conformational search.



INTRODUCTION

Proteins in solution are dynamic structures. Typically, loops are the regions with high mobility because they often contain substrate- or ligand-binding sites, the mobility being thus required for functioning.[1,2] Multistate proteins such as, e.g., molecular chaperones,[3,4] as well as intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs)[5−7] constitute an important part of every organism's proteosome. Therefore, instead of a fixed structure of a protein in solution, its dynamic ensemble should be considered.

Nuclear Magnetic Resonance (NMR) is a powerful technique for the determination of protein solution structures.[8] Protein-structure determination by NMR results in a bundle of conformations, consistent with the dynamic nature of proteins in solution. The conformations can be closely related, divergent in part, if the structure contains flexible regions, or even several alternative structural ensembles are obtained for multistate proteins. NMR structure determination is based on the conformational search with a given force field subject to experimental restraints, of which the distance restraints (usually between protons) or those imposed on dihedral angles are the most common.

The nature of NMR measurements implies that the restraints correspond to time- and ensemble-averaged quantities.[9,10] The first feature arises from a millisecond-scale mixing time of measuring the Nuclear Overhauser Effect that is the main source of distance restraints, while the second one from averaging over the whole solution ensemble. The determination of the structures of bioactive peptides by using time-averaged distance and angular restraints has a long history[11−19] but was not applied to proteins except for small ones.[20] Ensemble averaging can be performed in two manners: by including the restraints at simulation time through replica averaging[21−24] or by post-simulation reweighting of the resulting conformational ensemble.[25−29] Ensemble reweighting has found more practical applications and has been implemented, e.g., in the Xplor-NIH package.[28]

The extent and speed of conformational search, which is crucial for modeling the conformational ensembles of flexible proteins, can be increased tremendously with coarse-grained models, in which atoms are merged into extended interaction sites.[30−33] Apart from reducing the number of interaction sites, their advantage is the dilatation of the apparent time scale with respect to the all-atom or laboratory time scale due to removing explicit solvent molecules (for implicit-solvent models) and internal friction. This dilatation can amount to 3 orders of magnitude for heavily coarse-grained models.[34,35] The reso-

Received: November 6, 2024
Revised: December 27, 2024
Accepted: January 14, 2025
Published: January 24, 2025

lution of coarse-grained models is lower compared to all-atom ones but still seems reasonable, especially for flexible proteins.

In this study we developed an improved version of the time-averaged-restraint algorithm proposed by Torda et al.[11] and Bonvin et al.[14] and implemented it in the molecular dynamics (MD) with the UNRES coarse-grained model of polypeptide chains.[36,37] Owing to its physics-based derivation, in particular to our recently developed theory of coarse-graining,[38] UNRES is able to model protein structures and dynamics with considerable accuracy.[37,39] We used our recently developed ESCASA algorithm[40] to estimate proton positions from coarse-grained geometry analytically. We found that the method developed in this work restores the conformational ensembles from which synthetic restraints were generated and, for multistate proteins and proteins with disordered regions, gives ensembles better satisfying the experimental restraints than the respective ensembles deposited in the Protein Data Bank (PDB).[41]

METHODS

UNRES Model of Polypeptide Chains and its Implementation. In the UNRES model,[36,37] a polypeptide chain is represented as the trace of α-carbon (C^α) atoms, which are not interaction sites, linked with backbone virtual bonds, with the peptide groups (p) located halfway between the consecutive C^α atoms and the side chains (SC) attached to the C^α atoms with virtual bonds, these two kinds of objects being the interaction sites, as shown in Figure 1. The peptide-group sites represent the

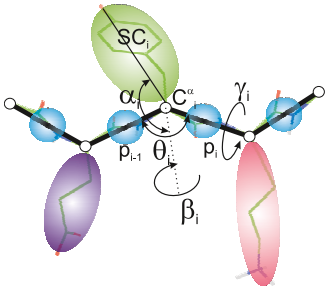


Figure 1. UNRES model of polypeptide chains. The interaction sites are united peptide groups located between the consecutive α-carbon atoms (light-blue spheres) and united side chains attached to the α-carbon atoms (spheroids with different colors and dimensions). The backbone geometry of the simplified polypeptide chain is defined by the $C^\alpha\cdots C^\alpha\cdots C^\alpha$ virtual-bond angles θ (θ_i has the vertex at C_i^α) and the $C^\alpha\cdots C^\alpha\cdots C^\alpha\cdots C^\alpha$ virtual-bond-dihedral angles γ (γ_i has the axis passing trough C_i^α and C_{i+1}^α). The local geometry of the ith side-chain center is defined by the zenith angle α_i (the angle between the bisector of the respective angle θ_i and the $C_i^\alpha\cdots SC_i$ vector) and the azimuth angle β_i (the angle of counterclockwise rotation of the $C_i^\alpha\cdots SC_i$ vector about the bisector from the $C_{i-1}^\alpha\cdots C_i^\alpha\cdots C_{i+1}^\alpha$ plane, starting from C_{i-1}^α). For illustration, the bonds of the all-atom chains, except for those to the hydrogen atoms connected with the carbon atoms, are superposed on the coarse-grained picture. Reproduced with permission from Zaborowski et al., J. Chem. Inf. Model., 55, 2050 (2015). Copyright 2015 American Chemical Society.

C′, O, N, and H backbone atoms and the side-chain sites represent the side-chain and the C^α and H^α atoms (Figure 1). The Cartesian coordinates of the C^α atoms and those of the SC centers are used as variables.[35] The energy function consists of site−site, local, and correlation potentials, the latter accounting for the coupling between the backbone-local and backbone-electrostatic interactions.[36−38,42] Most of the site−site interaction potentials have the axial and not the spherical symmetry. This feature and the presence of correlation terms are responsible for good performance of UNRES despite aggressive coarse graining. The UNRES effective energy function depends on temperature, which reflects the fact that it originates from the potential of mean force of polypeptide chains in water.[43] Details of the energy function and its derivation are described elsewhere.[36−38,42] In this work, we used the NEWCT-9P variant of the UNRES force field calibrated, by means of the maximum likelihood method, with a set of 9 proteins with different structural classes.[44]

The conformational search with UNRES is carried out with the use of MD, which has been implemented in our earlier work.[34,45] This implementation has further been extended[46] to multiplexed replica exchange molecular dynamics (MREMD),[47] with which the exploration of protein conformational space is more efficient. In MREMD, multiple trajectories are run at each temperature, which results in a more thorough search of the conformational space. To determine the weights of the conformations of an ensemble found by MREMD at the desired temperatures and, thereby, the ensemble averages, we use the binless variant of the weighted histogram analysis method (WHAM),[48] which was adapted to the temperature-dependent UNRES energy function in our earlier work.[49] The code has been parallelized in our earlier work[50] with the Message Passing Interface (MPI) libraries and, recently, heavily optimized and parallelized in the hybrid MPI/OpenMP mode.[35] Further, we ported the code to single[51] and multiple[52] Graphical Processor Units (GPUs).

Restraints from NMR with UNRES. Restraints are included as penalty terms added to the UNRES energy function. In this study, we imposed restraints on the distances between protons of different residues, on the $C^\alpha\cdots C^\alpha\cdots C^\alpha$ backbone-virtual-bond angles θ and on the $C^\alpha\cdots C^\alpha\cdots C^\alpha\cdots C^\alpha$ backbone virtual-bond dihedral angles γ. The θ and γ angles are shown in Figure 1. The proton coordinates to compute the distances are estimated analytically from the coarse-grained geometry by means of the ESCASA algorithm.[40,53] The angular restraints are derived from those on the backbone ϕ and ψ angles, by using eqs 10 and 22 from the paper by Nishikawa et al.[54] The extended energy function, including the penalty terms, is given by eq 1.

$$U = U_{\mathrm{UNRES}} + V_{\mathrm{dist}} + V_\theta + V_\gamma \tag{1}$$

where U_{UNRES} is the UNRES energy function, V_{dist} is the interproton-distance penalty term, and V_θ and V_γ are the respective angular penalty terms. These restraint potentials are given by eqs 2 − 4, respectively.[55−57]

$$V_{\mathrm{dist}}(d, d_{\mathrm{l}}, d_{\mathrm{u}}, A) = \begin{cases} A\dfrac{(d-d_{\mathrm{l}})^4}{\sigma^4+(d-d_{\mathrm{l}})^4}[1+\kappa \ln\cos h(d-d_{\mathrm{l}})] & \text{for } d < d_{\mathrm{l}} \\ 0 & \text{for } d_{\mathrm{l}} \le d \le d_{\mathrm{u}} \\ A\dfrac{(d-d_{\mathrm{u}})^4}{\sigma^4+(d-d_{\mathrm{u}})^4}[1+\kappa \ln\cos h(d-d_{\mathrm{u}})] & \text{for } d > d_{\mathrm{u}} \end{cases} \tag{2}$$

where d is a proton−proton distance (estimated from the coarse-grained coordinates), d_{l} and d_{u} are the lower and upper distance boundaries, respectively, which are taken from NMR data, σ is the thickness of the restraint-well wall, A is the depth of

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

the restraint-potential well, and κ is the slope of the restraint potential at large distances.[57] As in our earlier work,[53] we set κ = 0.01 to provide a small gradient driving at the desired distances but not to force the fulfillment of all restraints, which are likely to be contradictory for a system that consists of many interconverting conformations. We set σ = 1 Å, while A varied depending on the required restraint strength. It should be noted that σ determines the size of the attractor region of the penalty function. The value of σ established in our earlier work,[53] in which the restraints were not time averaged, was σ = 0.5 Å. However, in this work which concerns time-averaged restraints, we found that the attractor region should be larger and therefore increased σ.

$$V_\theta = A_\theta g(\theta, \theta_l, \theta_u) \tag{3}$$

$$V_\gamma = A_\gamma g(\gamma, \gamma_l, \gamma_u) \tag{4}$$

with

$$g(x, x_l, x_u) = \begin{cases} \frac{1}{4}\delta^4 & \text{for } \delta < \frac{x_l - x_u}{2} \\ 0 & \text{for } \frac{x_l - x_u}{2} < \delta < \frac{x_u - x_l}{2} \\ \frac{1}{4}\delta^4 & \text{for } \delta > \frac{x_u - x_l}{2} \end{cases} \tag{5}$$

$$\delta = \left(x - \frac{x_l + x_u}{2}\right) \bmod 2\pi \tag{6}$$

where θ_l, θ_u, γ_l, and γ_u are the lower and upper boundaries on the virtual-bond angles θ and virtual-bond-dihedral angles γ, respectively (which are calculated from the boundaries on the ϕ and ψ backbone dihedral angles). A_θ = 1 kcal/mol in this work and A_γ = 5 kcal/mol in this work are the restraint-potential-well depths.

The distances and angles in the penalty functions can be calculated from a single conformation or be time- or replica-averaged. In this work, we consider time-averaged restraints, which are described in section "Time-Averaged Restraints".

In our earlier work[53] we extended the distance-penalty term to treat ambiguous NOE restraints. This feature was not used in the current work because all assignments were unambiguous.

Time-Averaged Restraints. In the time-average-restraint method, the distances and angles from the simulated structures present in eqs 2−4 are averaged over a memory window[11−19] with an exponential memory function, as expressed by eq 7.

$$\overline{y}_j(t) = \left\{\left[\tau\left(1 - \exp\left(-\frac{t}{\tau}\right)\right)\right]^{-1} \int_0^t \exp\left(-\frac{t'}{\tau}\right) [y_j(\mathbf{r}(t - t'))]^{-m} dt'\right\}^{-1/m} \tag{7}$$

where $y_j(\mathbf{r}(t))$ is the jth restrained quantity, which depends on the coordinates that describe the geometry of the system (C^α atoms and side-chain centers in this work) collected in vector $\mathbf{r}(t)$ at time t of the trajectory, τ is the length of the memory window, and m is the exponent in averaging; m = 3 for distances and m = −1 for angles, the latter corresponding to direct averaging.[11,14,15,58] It should be noted that, except for small t, the normalization factor in eq 7 can be replaced with $[\tau(1 - e^{-1})]^{-1}$.

To evaluate the integral of eq 7, $Y_j(t)$, we use a recursive variant of the trapezoid formula (eq 8), which was also used in an earlier AMBER-package implementation of the time-averaged restraints.[58] In our implementation, the integral is permanently updated every n_{ave} time steps with the simple average $\langle Y_j \rangle$ over the time steps from $t_{n_{ave}I+1}$ to $t_{n_{ave}(I+1)}$, where $I = i \div n_{ave}$ is the number of n_{ave}-long sequences of MD steps until t_i. Between $t = t_{n_{ave}I+1}$ and $t = t_{n_{ave}(I+1)}$, Y_j is computed using the momentary value of y_j^{-m} (eqs 9 and 10). In this way, the target function always depends on the current molecular geometry and the restraint energy does not depend on trajectory history between $t = t_{n_{ave}I+1}$ and $t = t_{n_{ave}(I+1)}$, which assures total-energy conservation in this time period, should the simulations be run in the microcanonical mode. We demonstrate the advantage of this n_{ave}-time-step update of the integral with respect to every-time-step update in section "Stability of Time-Averaged Simulations".

$$Y_j(t_i) = \exp\left(-\frac{n_{ave}\Delta t}{\tau}\right) Y_j(t_{n_{ave}I}) + \left[\exp\left(-\frac{n_{ave}\Delta t}{\tau}\right)\langle Y_j(t_{n_{ave}I})\rangle + \Psi_j(t_i)\right] \frac{n_{ave}\Delta t}{2},$$
$$n_{ave}I < i \le n_{ave}(I + 1),\ I = i \div n_{ave} \tag{8}$$

$$\Psi_j(t_i) = \begin{cases} [y_j(t_i)]^{-m} & t_{n_{ave}I} < t_i < t_{n_{ave}(I+1)} \\ \langle Y_j(t_{n_{ave}(I+1)})\rangle & t_i = t_{n_{ave}(I+1)} \end{cases} \tag{9}$$

$$\langle Y_j(t_{n_{ave}(I+1)})\rangle = \frac{1}{n_{ave}} \sum_{i=n_{ave}I+1}^{n_{ave}(I+1)} [y_j(t_i)]^{-m} \tag{10}$$

We also set

$$Y_j(t_0) = [y_j(t_0)]^{-m} \tag{11}$$

Finally, the jth average quantity at time step t_i, $\overline{y}_j(t_i)$, is calculated from the integral of the $(-m)$th power, $Y_j(t_I)$, of the momentary observable $y_j(t) \equiv y_j(\mathbf{r}(t))$ over the trajectory until the t_ith time step (eq 12).

$$\overline{y}_j(t_i) = \left[\frac{Y_j(t_i)}{\tau(1 - e^{-1})}\right]^{-1/m} \tag{12}$$

To calculate the time averages of the dihedral angles, their sines and cosines are time-averaged first and the average dihedral angles are calculated from these quantities by using the standard FORTRAN `atan2` function.

The gradient due to time-averaged restraints is expressed by eq 13.

$$\nabla_{r_k} V(t_i) = \sum_{j=1}^{N_r} \frac{\partial V(\overline{y}_j(t_i))}{\partial \overline{y}_j} \frac{\partial \overline{y}_j(t_i)}{\partial y_j} \nabla_{r_k} y_j(t_i) \tag{13}$$

where $V(t_i)$ is a shorthand for $V(\mathbf{r}(t_i))$, N_r is the total number of restraints of a given kind (distance or angular in this work), and

$$\frac{\partial \overline{y}_j(t_i)}{\partial y_j} = \frac{1}{2}\left[\frac{\overline{y}_j(t_i)}{y_j(t_i)}\right]^{m+1} \frac{n_{ave}\Delta t}{\tau(1 - e^{-1})} \tag{14}$$

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

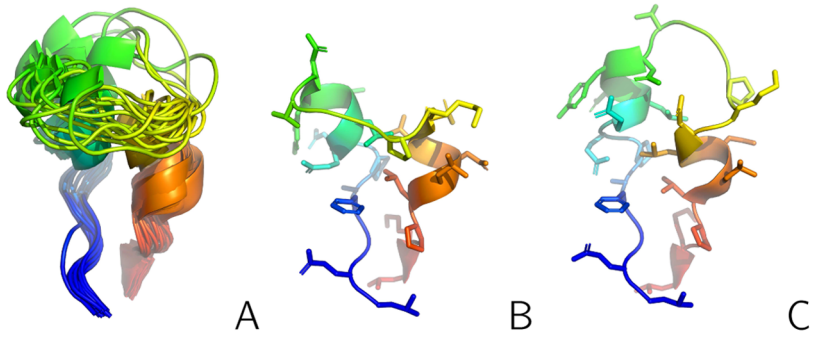


Figure 2. (A): The NMR-determined conformational ensemble of the L129−L153 fragment of 2KW5 from the respective PDB entry. (B, C): Structures #1 and #6 from this ensemble, which were used to calculate synthetic interproton-distance restraints. The backbone is shown in the cartoon representation, while the side chains in panels (B, C) are shown in the stick representation. The chains are colored from blue to red from the N- to the C-terminus. The side chains are omitted from panel (A). The drawings were made with PyMOL.[60]

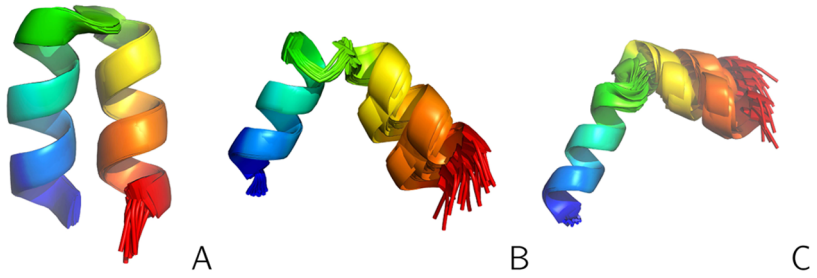


Figure 3. NMR-determined conformational ensembles of the three forms of 2LWA corresponding to structures A (panel A), B (panel B) and C (panel C)[61] of the respective PDB entry. The backbones are shown in the cartoon representation and the side chains are omitted. The chains are colored from blue to red from the N- to the C-terminus. The drawings were made with PyMOL.[60]

From eq 14 it follows that, for $t_i \gg \tau$, the derivatives of the time-averaged quantities and, thereby, the time-averaged-restraint gradients (eq 13) are scaled down by the factor of $n_{ave}\Delta t/[\tau(1-e^{-1})]$. The values of τ usually applied range from 5 to 50 ps.[14,15] In this work we set τ from 4.89 to 489 ps (in UNRES MD, a "natural" time unit equal to about 48.9 fs is applied,[45] hence we set τ to its integer multiplicity). Therefore, especially with $n_{ave} = 1$, the restraint contributions to forces are reduced to effectively turn the MD runs with time-averaged restraints into unrestrained runs. To ensure large enough force contributions from time-averaged restraints, we scale the time-averaged-restraint gradients by a factor f_i (at the ith time step) defined by eq 15.

$$f_i = \begin{cases} 1 & i = 0 \\ (1-e^{-1})\dfrac{i\tau}{N_\tau n_{ave}\Delta t} & i \le N_\tau \\ (1-e^{-1})\dfrac{\tau}{n_{ave}\Delta t} & i > N_\tau \end{cases} \quad (15)$$

where N_τ is the number of MD steps over which the scaling factor should be increased from 1 to $(1-e^{-1})\tau/(n_{ave}\Delta t)$. This parameter is typically equal to the integer part of $\tau/\Delta t$ but can be set by the user.

The gradual scaling of the restraint forces with the progress of an MD trajectory prevents us from overemphasizing the restraint contributions until enough MD steps have been executed to provide at least $\tau/(n_{ave}\Delta t)$ terms in the averaged quantities. On the other hand, we determined that the time-averaged-restraint components of the potential energy, which come into play when REMD/MREMD simulations are run, should not be scaled or the energy differences between replicas become too big for any replica exchange to happen. Moreover, postprocessing the REMD/MREMD simulations with WHAM to determine the weights of the conformations encounters insurmountable numerical problems due to big differences in the values of the scaled time-average-restraint energy components. We will demonstrate the advantages of the revised time-averaged-restraint algorithm proposed in this work in section "Essential Role of Restraint-Force Scaling in Time Averaged Simulations".

Systems Studied. To investigate the behavior of the UNRES MD/MREMD simulations with time-averaged restraints, we selected the L129−L153 loop part of the Slr1183 protein from *Synechocystis sp.* (PDB: 2KW5[59]). The respective conformational ensemble derived from the PDB: 2KW5 structure is shown in Figure 2A. It can be seen that the structure is largely disordered. For this system, we carried out calculations with synthetic interproton-distance restraints derived from structures #1 and #6 of the ensemble. These structures are shown in panels B and C of Figure 2. This system is hereafter referred to as 2KW5(129−153).

To test the performance of UNRES with the time-averaged NMR-derived restraints feature with multistate proteins, we selected the influenza hemagglutinin fusion peptide (PDB: 2LWA,[61] 25 residues). The NMR-determined conformational ensembles of three distinct forms of this protein (referred to in its PDB entry as chains A, B, and C, respectively, which are hereafter referred to as structures A, B, and C, respectively, in this paper), are shown in Figure 3A−C.

To test our method with larger-size (partially disordered and not multistate) proteins, we selected two proteins from the Montelione/NEF Benchmark Data Set,[62] for which both NMR and X-ray structures are available. These were the complete

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

structure of 2KW5 (202 residues; its X-ray structure counterpart being 3MER), hereafter referred to as 2KW5 and the peptide methionine sulfoxide reductase msrB from *Bacillus subtilis* [PDB: 2KZN[59] (NMR), 3E0O (X-ray), 147 residues], hereafter referred to as 2KZN. For reference, we included one more protein from the Montelione/NEF Benchmark set, which has a well-defined structure, namely the *Staphylococcus aureus* protein SAV1430 [PDB: 1PQX (NMR), 2FFM (X-ray), 91 residues], hereafter referred to as 1PQX. The NMR structures of 2KW5 and 2KZN contain quite large disordered segments, which was reflected in the results of our earlier work,[53] in which we found that the structures deposited in the PDB are consistent only with about 50% of experimental distance restraints. The NMR structure of 1PQX (which is well-defined) satisfies almost all of the experimental restraints. The NMR ensembles and the X-ray structures of these proteins are shown in Figure 4A–C. As in our previous work,[53] we used the X-ray structures of these three proteins as reference structures.

Calculation Procedure. Most of the canonical MD and MREMD simulations were carried out in the Langevin-dynamics mode (which also provides thermostatting), with the time step of Δt = 4.89 fs. In some of the runs, which were aimed at checking the conservation of the bath temperature, the Berendsen thermostat[63] was also used. The velocity-Verlet integrator[64] adapted to the Langevin dynamics with UNRES,[34] which uses the variable-time-step (VTS) algorithm[45] was applied to solve the equations of motion. All simulations were carried out with the recently optimized UNRES code.[35]

All starting structures were random-generated by using the procedure described in ref 65. Briefly, a coarse-grained polypeptide chain is gradually built up starting from the N-terminus. To add the next residue, its backbone-virtual-bond-dihedral angle γ is sampled at random from the $[-180°, 180°]$ interval, while its backbone-virtual-bond-angle θ as well as the zenith (α) and the azimuth (β) angles defining the local geometry of united side chain (Figure 1) are sampled from the Boltzmann distributions calculated from the respective knowledge-based potentials determined in ref 66. The Cartesian coordinates of the residue are calculated from the generated internal coordinates. Subsequently, its overlaps with the previous residues are checked and, if found, the generation is retried. If 100 retrials are unsuccessful, the generation is restarted from one more residue backward. For multitrajectory canonical simulations and MREMD simulations, different starting structures were generated for different trajectories.

For 2KW5(129–153), we ran both canonical and micro-canonical simulations, the latter to determine the extent of energy conservation. Each series of canonical simulations was run at T = 300 K and consisted of 4 or 8 trajectories, 5,000,000 to 10,000,000 MD steps each.

For 2LWA, 2KW5, 2KZN, and 1PQX, MREMD simulations only were run at the following 12 temperatures: 260, 262, 266, 271, 276, 282, 288, 296, 304, 315, 333, and 370 K, respectively, which were selected by using the Hansmann algorithm[67] to maximize the walks in the temperature space. Four replicas were run at a given temperature, resulting in a total of 48 replicas. Each replica consisted of 10,000,000 (2LWA) or 20,000,000 (the other proteins) time steps and the temperatures were exchanged between replicas every 10,000 time steps. The temperature was controlled by the Langevin thermostat, with scaling down the water friction by a factor of 0.05. The UNRES coordinates were saved every 10,000 time steps, i.e., every replica-exchange time. The last 1000 structures from each trajectory (48,000 structures total) were taken for further analysis.

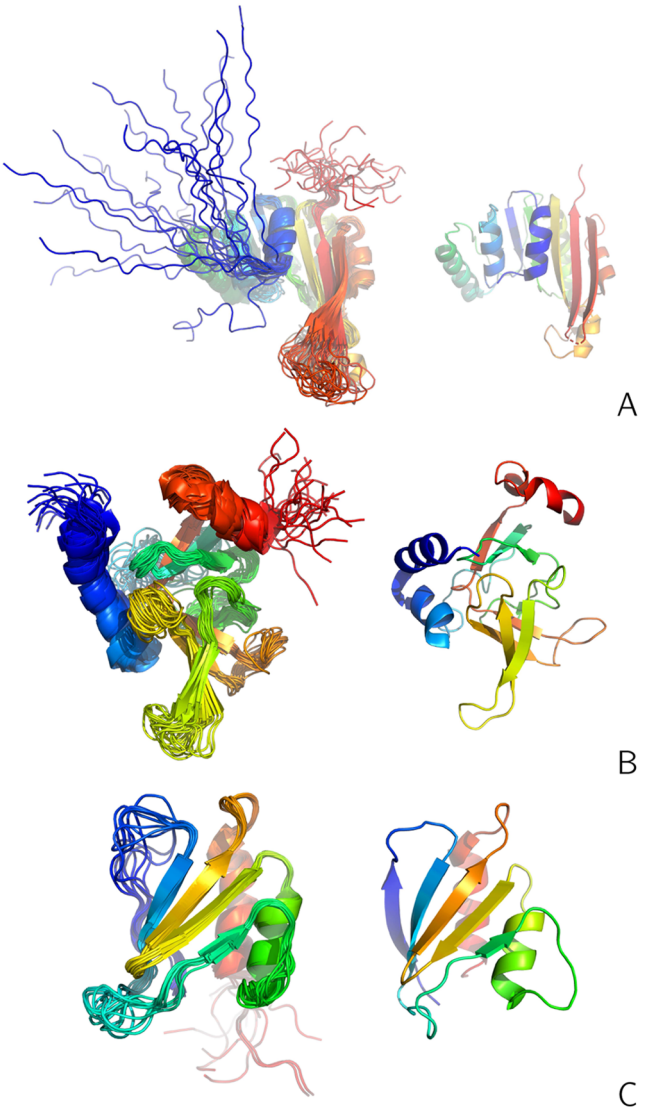


Figure 4. NMR-determined conformational ensembles (left) and the X-ray structures (right) of the three proteins of the Montelione/NEF Benchmark Data Set[62] considered in this study: (A) 2KW5 (NMR)/3MER (X-ray); (B) 2KZN (NMR)/3E0O (X-ray); (C) 1PQX (NMR)/2FFM (X-ray). The backbones are shown in the cartoon representation, the chains colored from blue to red from the N- to the C-terminus. The drawings were made with PyMOL.[60]

The structures resulting from MREMD simulations were subjected to postprocessing with the UNRES implementation[49] of the binless variant of WHAM[48] to enable us to compute the statistical weights at the desired temperature within the replica-temperature range. Because the aim of time-averaged MREMD simulations is to produce equilibrium ensembles subject to time-averaged restraints, the time-averaged penalty terms are present in the extended effective energy expression (eq 1) which appear in the WHAM equations (eqs 14–16 in ref 49). The part of an ensemble comprising 99% of conformations at a given temperature was subjected to a cluster analysis by means of the Ward minimum variance method.[68] In this work, we collected ensembles at T = 280 K (for computing the ensemble-averaged distances and constructing the PDB-entry-style sets of NMR-determined conformations) or T = 260 K and T = 280 K

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476–1493

for determining single conformations to compare with the X-ray structures. For comparison with the reference X-ray structures, the number of clusters (and, thereby, the number of models) was set at 5, this number being selected after the rules of Community Wide Experiments on the Critical Assessment of Techniques for Protein Structure Prediction (CASP),[69] in which 5 models per target can be submitted for assessment. The families (and, consequently, the selected structures) were ranked by the cumulative probabilities of all conformations belonging to them, as described in our earlier work.[49] The structure with the lowest restraint violation was selected as the representative of a given family. Additionally, the ensembles of 2LWA, 2KW5, 2KZN, and 1PQX were dissected into 20 families whose representative conformations best fitting the NMR data were selected to constitute reduced ensembles with a size typical of NMR-determined ensembles deposited in the PDB.

To compute the actual (not estimated with ESCASA) interproton distances and, subsequently, average interproton distances at the all-atom level, we ranked all structures of a given system according to decreasing weights determined by WHAM, given T = 280 K, and took the ensemble composed of those whose sum of weights was 0.99. These structures were subsequently converted to all-atom representation by using the cg2all algorithm,[70,71] followed by refinement with AMBER[72] with the ff19SB force field[73] and implicit-solvent Generalized Born Surface Area (GBSA) model,[74,75] as described in our earlier work.[56] Then we computed the interproton distances for all structures. Finally, we computed the r^{-6}-averaged distances using eq 16.

$$\bar{d}_{H_i(k)H_j(k)} = \left[\sum_{l=1}^{N_e} w_l d^{-6}_{H_i(k)H_j(k);l}\right]^{-1/6} \tag{16}$$

where $d_{H_i(k)H_j(k)}$ is the average distance between the protons that belong to restraint k, N_e is the number of structures in the ensemble, w_l is the weight of the lth conformation of the ensemble (determined by WHAM; the weights are normalized to 1), and $d_{H_i(k)H_j(k);l}$ is the interproton distance for the lth conformation. If the restraint corresponds to an average over equivalent proton groups (such as, e.g., the H^β protons of isoleucine or the H^δ protons of phenylalanine), all distances between the individual protons of these groups are r^{-6}-averaged for a given conformations, with equal weights.

The same procedure was applied to the reduced ensembles of conformations of 2LWA, 2KW5, 2KZN, and 1PQX (for which the weights were summed over each cluster to obtain the weight of the representative of a given family representative) and to the PDB ensembles (for which the weights of all conformations were equal).

The distance-boundary violations were quantified as the right upper distance boundary root-mean-square deviations (ρ_u^+) defined by eq 17, the total number of interproton distances greater than the respective upper distance boundary, and the number of distances greater by 2 Å or more than the upper distance boundaries.

$$\rho_u^+ = \sqrt{\frac{1}{N_d}\sum_{i=1}^{N_d}\delta_i^2} \tag{17}$$

$$\delta_i = \begin{cases} \bar{d}_i - d_i^u & d_i > d_i^u \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where N_d is the number of interproton-distance restraints, $\bar{d}_i$ is the ensemble-averaged (cf. eq 16) distance from the simulated or PDB ensemble and d_i^u is the upper distance boundary.

As measures of fitting the simulated structures to the reference structures, we used the root-mean-square deviation of the C^α-atom positions (C^α-RMSD or, in short, RMSD) at the optimal superposition of the C^α atoms of the target structure on those of the reference structure and the Global Distance Test Test Score (GDT_TS).[76,77] The latter is an average of the percentages of the target structure whose C^α atoms superpose within 1, 2, 4, and 8 Å, respectively, on those of the reference structure. The TMScore software from Zhang's lab[78] was used to calculate both measures.

Experimental and Synthetic Restrains. The experimental restraints were the version-2 (v.2) restraints taken directly from the PDB entries and converted to the UNRES input format. The distance restraints pertaining to the same residue were omitted because they do not contribute useful information at the coarse-grained level.

For testing the method with the 2KW5(129−153) system, we generated interproton-distance restraints from conformation #1 (Figure 2B; a total of 173 restraints) and conformation #6 (Figure 2C; a total of 159 restraints), as well as r^{-6}-averaged distance restraints from both structures, calculated by using eq 16 with two weights only, each one equal to 0.5, of its PDB ensemble (a total of 186 restraints). These distance restraints are collected in the respective machine-readable files of the `Suppdata.zip` archive of the Supporting Information (see section "Glossary of the machine-readable files" of the Supporting Information for detailed content of the archive).

The experimental distance and angular restraints (both the original restraints on the ϕ and ψ angles and those on the θ and γ angles after transformation to the coarse-grained representation, which were used in the actual calculations) for 2LWA, 2KW5, 2KZN, and 1PQX are collected in the respective machine-readable files of the `Suppdata.zip` archive of the Supporting Information and the numbers of these restraints are collected in Table 1. Although both distance and angular restraints were used in the calculations, in what follows we discuss only the violations of the distance restraints. The reason for this is that even if the angular restraints on the coarse-grained geometry are fulfilled, the original restraints on the ϕ and ψ angles (the values of which are calculated after converting the coarse-grained structures to the all-atom structures) are often violated. Therefore, restraints should be imposed on the ϕ and ψ angles estimated analytically from the coarse-grained geometry in a similar way as the estimation of proton positions in the ESCASA approach.[40] In this work, we treat the restraints on the θ and γ angles derived from those on the ϕ and ψ angles only as a crude way to keep control over the backbone-local geometry.

Table 1. Numbers of NMR-derived Restraints on Interproton Distances, Backbone-virtual-bond-dihedral Angles γ, and Backbone-virtual-bond Angles θ Used in NMR-data-assisted UNRES/MREMD Simulations

	# restraints on		
protein	distances	γ	θ
2LWA	175	18	20
2KW5	947	134	147
2KZN	578	104	121
1PQX	1015	64	74

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

RESULTS AND DISCUSSION

Stability of Time-Averaged Simulations. Because the potential-energy function in time-averaged-restraint simulations contains a time-dependent term (eq 7), the total energy is not conserved in the microcanonical mode. To determine the extent of energy nonconservation in time-averaged-restraint simulations, we carried out MD runs in the microcanonical (NVE) mode (constant number of particles, volume and total energy) for the 2KW5(129−153) system with the synthetic distance restraints generated from conformations #1 and #6 of its NMR ensemble (see the Supporting Information for the name and location of the restrain files) without (run 1) and with time averaging (runs 2−4), respectively. For each run, 100,000 MD steps were executed with a small time step of Δt = 0.489 fs, the initial velocities corresponding to the temperature of 300 K (however, the temperature is not conserved in NVE runs). The value of τ in eq 7 in runs 2−4 was 4.89 ps and the value of n_{ave} was 1 in runs 2 and 3 and 100 in run 4. The restraint-potential well-depth (A in eq 2) was 5 kcal/mol. In run 2, the restraint forces were not scaled, while in runs 3 and 4 they were scaled up by the factor of $(1 - e^{-1})\tau/(n_{ave}\Delta t)$ (cf. section "Time-Averaged Restraints"). The plots of the total energy vs time are shown in Figure 5A,B. It can be seen from Figure 5A that the total energy remains effectively constant (the oscillations not exceeding 0.001 kcal/mol) in restrained microcanonical MD simulations without time averaging (run 1). In all time-averaged simulations (runs 2−4), the total energy is not constant even when the time-averaged restraint forces are not scaled, which introduces only a small perturbation. The initial total-energy drift observed in the plots arises because the 4.89 ps memory window is sufficiently filled only after about 10,000 MD steps, during which the averaged restraints are being built, but the total energy continues to vary after this. The average-update frequency (n_{ave}) does not have a major effect on the total-energy variation in a long run. However, as can be seen from Figure 5B, in which the total energy is plotted for the last 1000 steps of runs 3 and 4, the total energy stays constant in the n_{ave}-step-long intervals.

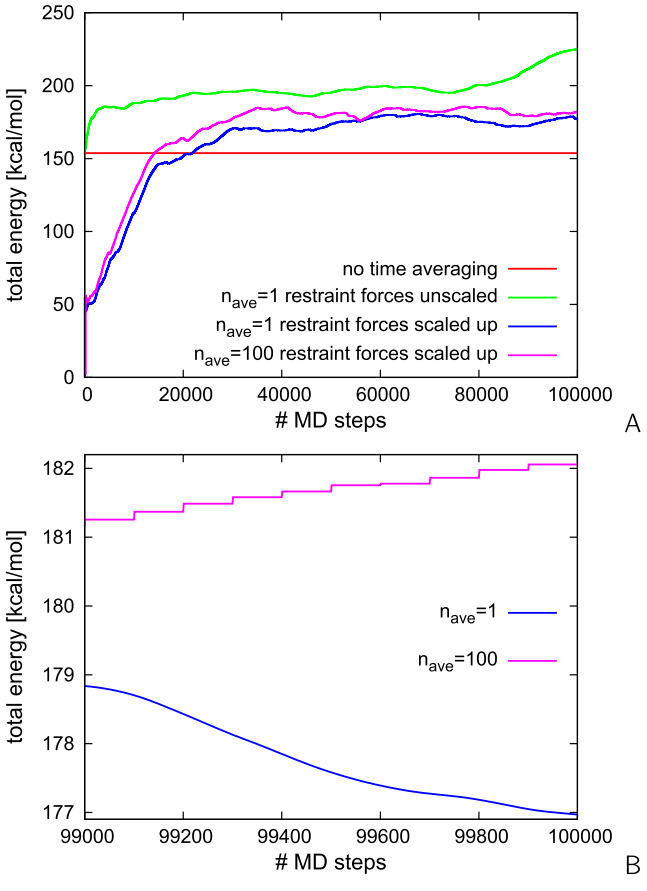


Figure 5. (A) Variation of the total energy in the restrained NVE MD (with a small time step of Δt = 0.489 fs) runs of 2KW5(129−153) without time averaging (run 1) and with time averaging (τ = 4.89 ps): no force scaling, n_{ave} = 1 (run 2), force scaling up by $(1 - e^{-1})\tau/\Delta t$: n_{ave} = 1 (run 3), and n_{ave} = 100 (run 4). (B) A close-up of run 3 and run 4 for the last 1000 time steps. In runs 3 and 4, the restraint energy was scaled up along with the restraint forces to check total-energy conservation. The plots were made with gnuplot.[79]

Nonconservation of the total energy in the microcanonical mode could result in a substantial increase of the kinetic temperature in the canonical mode and, thereby, in significant errors in the generated conformational ensembles. To determine the extent of the problem, we carried out a series of canonical (NVT) runs for the 2KW5(129−153) system (with restraint-potential well-depth A = 5 kcal/mol), each consisting of 4 independent 10,000,000-step trajectories with the time step of Δt = 4.89 fs (which is the time step used in the production simulations), without time averaging and with different variants of time averaging. The temperature was controlled by the Berendsen[63] or the Langevin thermostat, the latter with scaling the water friction by a factor of 0.02 or 0.05, respectively. The thermal-bath temperature was set at T_{bath} = 300 K. The average kinetic temperature was calculated from the second half of each of the 4 trajectories of a given run. The average temperatures, along with run settings, are collected in Table 2. It can be seen that the average kinetic temperature does not differ much from that of the thermal bath for the runs without time averaging but it is significantly higher in the time-averaged-restraint runs for the smallest value of τ = 4.89 ps, the difference decreasing with increasing n_{ave}. With the highest τ = 489 ps and with n_{ave} = 1000, the average temperatures approach those obtained for the runs without time averaging. The temperature is the most close to the bath temperature for the Berendsen thermostat; however, using this thermostat results in a very narrow kinetic-energy distribution. Therefore, we selected the Langevin thermostat

Table 2. Average Kinetic Temperature from the Series of 4-Trajectory MD Runs (10,000,000 Steps with a 4.89 fs Time Step Per Trajectory) without Time Averaging and with Time Averaging with Different Values of τ and n_{ave}

τ [ps]	n_{ave}	T_{kin} [K] Ber.	Lang. 0.02	Lang. 0.05
No Time Averaging				
		300.0	301.4	303.1
Time Averaging				
4.89	1	327.2	542.1	365.2
	10	326.6	532.1	362.1
	100	316.0	425.9	336.2
48.9	1	306.0	346.0	318.4
	10	306.0	345.1	318.4
	100	305.4	341.1	317.2
	500	303.5	328.2	312.7
	1000	302.2	318.9	309.4
489.0	1000	300.5	305.3	304.6

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

with water-friction scaling of 0.05, even though it results in a slightly slower dynamics compared to that with the scaling factor of 0.02. Based on the obtained results, we used τ = 48.9 ps and n_{ave} = 500 or τ = 489 ps and n_{ave} = 1000 in most of the calculations. We also noted that temperature conservation becomes better as the system size increases.

Essential Role of Restraint-Force Scaling in Time Averaged Simulations. To demonstrate the necessity of restraint-force scaling in restrained MD simulations with time averaging, we carried out 4 series of canonical-MD runs with 2KW5(129−153). Each run consisted of 4 trajectories, 1,000,000 steps each with a 4.89 fs time step, at T = 300 K. It can be seen from Table 2 that the kinetic temperature is for this system by about 36 deg higher than the bath temperature when the time-averaged-restrain forces are scaled. The rationale of setting a relatively small τ was to check how does the method perform at the boundary of stability limit. Runs 1−3 were carried out with restraints derived from structure #1. Run 1 was carried out without time averaging, while runs 2 and 3 were carried out in the time-averaged mode with τ = 4.89 ps, with (run 2) and without (run 3) restraint-force scaling, respectively. Finally, run 4 was carried out without any restraints.

The plots of the C^{α}-RMSD from structure #1 are shown in Figure 6. As can be seen, low RMSD has been obtained in the

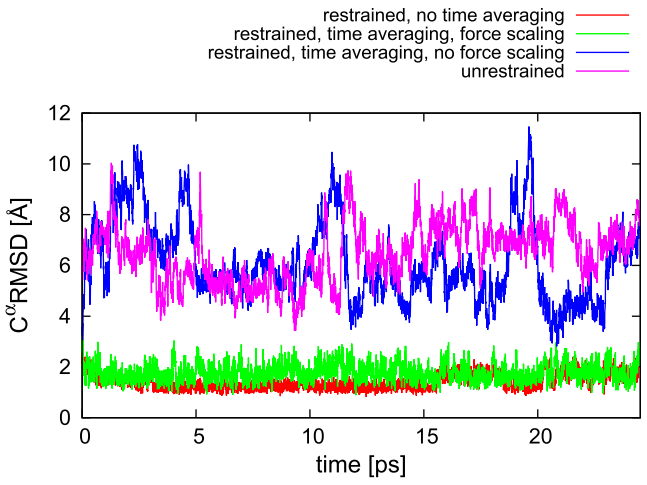


Figure 6. Superposed plots of the variation of C^{α}-RMSD of the conformations of 2KW5(129−153) from structure #1 obtained in restrained canonical MD simulations with UNRES (restraint well depth A = 5 kcal/mol) with interproton-distance restraints calculated from structure #1 without time averaging (run 1), with time averaging (τ = 4.89 ps, n_{ave} = 10) and restraint-force scaling (run 2), with time averaging and no restraint-force scaling (run 3), and in an unrestrained MD run (run 4). The plot was made with gnuplot.[79]

restrained simulations without time averaging. This result could be expected because the restraints were derived from a single conformation. The simulation with time averaging and without restraint-force scaling resulted in high RMSD values given the small size of the system considered here, the RMSD range being not much different from that from unrestrained simulations. Only with restraint-force scaling did the RMSD values become comparable to those from regular restrained simulations.

It should be noted that the kinetic temperature for the run with unscaled time-averaged restraint forces is 303.3 K, which means that the high RMSD obtained in this run could not result from the elevated kinetic temperature but only from not scaling the restraints. As pointed out above, the kinetic temperature was remarkably elevated with scaled restraint forces which, however, did not prevent the simulation from reaching conformations close to the reference structure. This result demonstrates the robustness of the method even when the time window is short, which results in a remarkably elevated kinetic temperature with respect to the bath temperature.

Effect of Restraint Averaging on Simulated Structures. In this part of our study, we carried out calculations with synthetic restraints derived from structure #1, #6, or both, of 2KW5(129−153) (Figure 2B,C), in order to determine the behavior of the method when the reference structure(s), which the restraints correspond to are known.

First, we determined the effect of time averaging on the results of the simulations with the restraints derived from one well-defined structure. To accomplish this task, we carried out canonical MD simulations of 2KW5(129−153) with synthetic interproton-distance restraints derived from structure #1. We carried out one series of runs with nonaveraged restraints, with the restraint-potential well-depth A = 1 kcal/mol, and two series of runs with time-averaged restraints, with τ = 4.89 ps, n_{ave} = 100 and A = 1 or 5 kcal/mol, respectively. The relatively small τ was set to determine if the method can produce conformational ensembles compatible with the restraints even with a small memory-window length. Each series consisted of 4 trajectories of 10,000,000 MD steps with a 4.89 fs step size. The plots of the RMSD distribution functions collected from the last 2,000,000 steps of all trajectories of a given series are shown in Figure 7. As

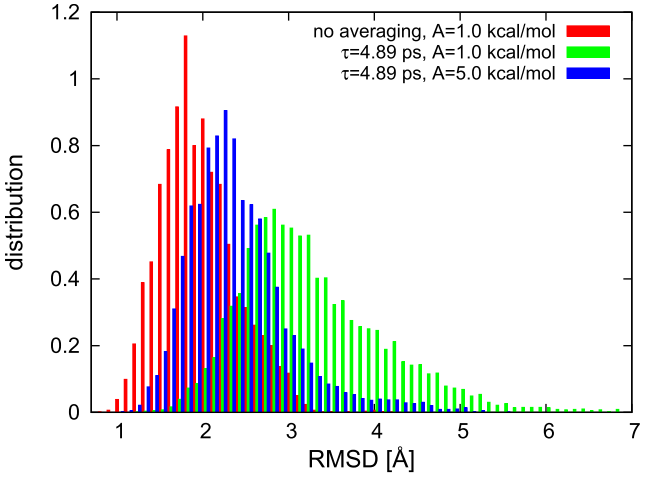


Figure 7. C^{α}-RMSD distributions from the NMR structure #1 of the 2KW5(129−153) conformations obtained in canonical MD simulations with interproton-distance restraints derived from structure #1 without time averaging and with time averaging with the restraint-well depth of 1.0 and 5.0 kcal/mol, respectively. See text for the other settings of the simulations. The plots were made with gnuplot.[79]

shown, introducing time-averaged restraints results in right-shifting and broadening the RMSD distribution, increasing the restraint well-depth to 5 kcal/mol resulting in a smaller shift. Increasing A beyond the value of 5 kcal/mol reduces the shift but we found that already the value of 10 kcal/mol could cause instability due to excessive time-average-restraint forces. Consequently, the value of 5 kcal/mol is a reasonable compromise between approaching the results of restrained simulations without time averaging for nonaveraged restraints and numerical stability. The shift results from time averaging, owing to which the distances are averaged over a time window

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

and, consequently, a single conformation does not need to satisfy all restraints. Therefore, the obtained ensemble is more diffuse compared to that resulting from simulations without time averaging. Because the RMSD range has the left boundary of 0, the distribution is asymmetric, as can be seen from Figure 7. Consequently, increasing the spread of the conformational ensemble not only makes the RMSD distribution broader but also right-shifted.

Subsequently, we carried out the canonical MD simulations for the 2KW5(129−153) system with the synthetic restraints derived from structures #1 and #6, without (run series 1) and with (run series 2) time averaging, respectively. For reference, we carried out two series of runs without and with time averaging with restraints derived from structure #1 (run series 3 and 4) and two series of runs with restraints derived from NMR structure #6 only (run series 5 and 6). Each run series consisted of 4 trajectories, 10,000,000 4.89 fs time steps each, with the restraint-potential-well depth A = 5 kcal/mol.

The variation of C^{α}-RMSD from structures #1 and #6 for two sample trajectories from run series 1 is shown in Figure 8A,B and that for a sample trajectory from run series 2 is shown in Figure 8C, respectively. The other trajectories exhibit similar patterns of RMSD variation. As can be seen, without time averaging, the resulting structures are either stuck in the neighborhood of structure #1 (low RMSD from structure #1 and high from structure #6; Figure 8A) or in conformations with a moderate distance from structure #1 and structure #6 (Figure 8B). Conversely, with time averaging, the system alternates between the two parent structures (Figure 8C).

The two-dimensional probability-distribution maps in the RMSDs from NMR structures #1 and #6, respectively ($RMSD_1$ and $RMSD_6$, respectively) for all six run series are shown in Figure 9A−F. It can be seen that the RMSD distribution obtained from run series 1 (Figure 9A) indicates that the obtained ensemble is closest to NMR structure #1, with only a minor part of structures being kind of similar to structure #6. Conversely, with time averaging (run series 2), there are two clear lobes, one corresponding to conformations closer to structure #1 and the other one to structure #6, respectively (Figure 9B). The distributions from run series 3 and 4 (Figure 9C,D) demonstrate that the restraints from structure #1 exclusively result in conformations closer to structure #1 than those obtained with restraints averaged over structures #1 and #6. This observation pertains to the structures obtained both with and without time averaging, the RMSD being lower for the run series without time averaging, as already observed in Figure 7. A similar conclusion regarding the similarity to structure #6 can be drawn by comparing the distributions from run series 2 (Figure 9B) and run series 5 and 6 (Figure 9E,F). However, with the restraints from structure #6 and without time averaging, the RMSD distribution has 3 lobes, only one of which corresponds to conformations close to structure #6 (Figure 9E). Conversely, with time averaging, the RMSD distribution is unimodal, with the maximum corresponding to conformations close to structure #6 and far from structure #1. Consequently, while time averaging shifts the distribution of conformations slightly farther away from the reference structure, it seems to improve the ergodicity of simulations.

Test with a Small Multistate Protein (2LWA). To test the performance of the data-assisted protein-structure modeling with UNRES and time-averaged restraints, we selected the 2LWA protein. The original paper on its NMR structure determination with the aid of Xplor-NIH[61] reports three families

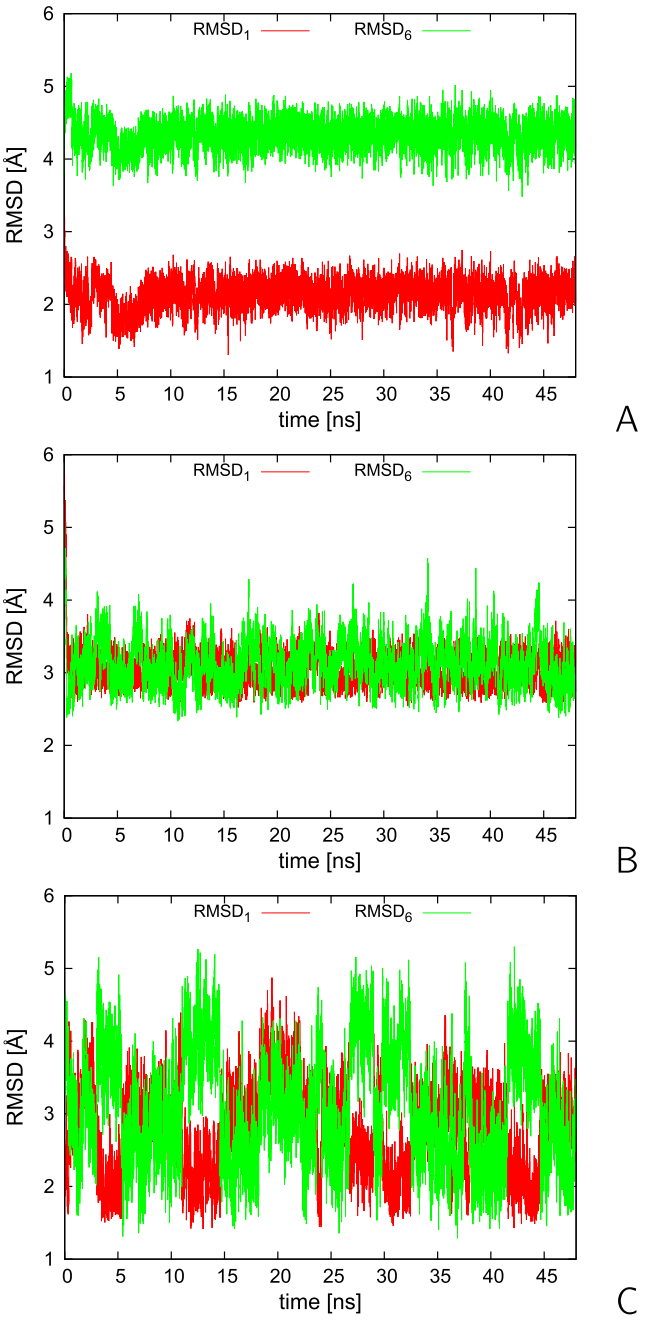


Figure 8. Variation of C^{α}-RMSD from the PDB-ensemble structures #1 ($RMSD_1$) and #6 ($RMSD_6$) of the 2KW5(129−153) system with simulation time in sample trajectories of canonical MD simulations with synthetic average interproton-distance restraints derived from structures #1 and #6. (A, B) Simulations without time averaging. (C) Simulations with time averaging (τ = 4.89 ps, n_{ave} = 10). The plots were made with gnuplot.[79]

of solution conformations: a helical hairpin (structure A of the 2LWA PDB entry), a partially open helical hairpin (structure B) and a kinked helix (structure C), as shown in Figure 3. We ran the NMR-assisted UNRES/MREMD simulations at 12 temperatures, each quadruplexed (48 replicas total), the temperatures distributed as described in section "Calculation Procedure". We implemented both the distance restraints (with the restraint-well depth A = 5 kcal/mol; eq 2) and the restraints on the backbone-virtual-bond angles θ (eq 3) and backbone-virtual-bond-dihedral angles γ (eq 4) calculated from the original restraints

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

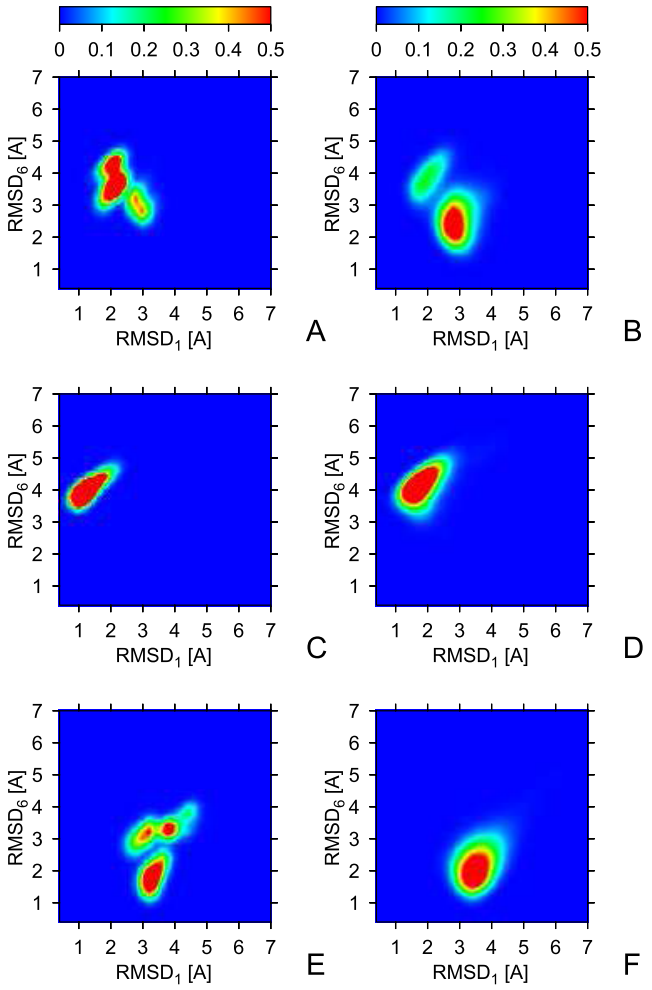


Figure 9. Two-dimensional maps of the C^{α}-RMSD distributions from structure #1 ($RMSD_1$) and #6 ($RMSD_6$) calculated from the 2KW5(129−153) conformations obtained in canonical MD simulations with interproton-distance restraints derived from structures #1 or #6 or both. (A) Simulations without time averaging, restraints from structures #1 and #6. (B) Simulations with time averaging, restraints from structures #1 and #6. (C) Simulations without time averaging, restraints from structure #1. (D) Simulations with time averaging, restraints from structure #1. (E) Simulations without time averaging, restraints from structure #6. (F) Simulations with time averaging, restraints from structure #6. See text for the other settings of the simulations. The color scale of probability density is above the graphs. The plots were made with GRI.[80]

on the ϕ and ψ backbone-dihedral angles. The time-averaged simulations were carried out with τ = 48.9 ps, n_{ave} = 500 and τ = 489 ps, n_{ave} = 1000. The latter gave slightly better results in terms of fitting the calculated ensemble-averaged interproton distances to the NMR data.

Subsequently, for each run, we constructed a representative subensemble of 20 conformations, as described in section "Calculation Procedure". The PDB files of these ensembles are included in the `Suppinfo.zip` archive of the Supporting Information. The ensembles obtained without and with time averaging, with τ = 489 ps, n_{ave} = 1000, are shown in the C^{α}-trace representation in Figure 10A,B, respectively. For reference, the ensemble composed of all superposed structures from the 2LWA PDB entry is shown in Figure 10C. It can be seen that all structures obtained without time averaging are helical hairpins forming a tight bundle and are thus similar to the 2LWA structure A but not to structures B and C deposited in the PDB (cf. Figure 3A). With time averaging, the obtained structures correspond to helical hairpins with the gradually increasing interhelix angle (Figure 10B). This feature of the obtained ensemble is emphasized in panel D of the Figure, in which three representative conformations are shown in the cartoon (backbone) and stick (side chains) representation. These structures are similar to structures A, B, and C, respectively, of the 2LWA PDB entry. Compared to the ensemble of 2LWA obtained with time averaging (Figure 10B), which consists of a continuity of conformations, the compact conformations of structure A from the 2LWA PDB entry are clearly distinguished from those of structures B and C.

The differences between the ensembles resulting from the two modes of restrained simulations are also illustrated in Figure 11A,B, where the C^{α}-RMSD distributions from the PDB structures A, B, and C are plotted. As can be seen from Figure 11A, narrow RMSD distributions are obtained without time averaging and that corresponding to structure A is remarkably shifted to the left. The lowest RMSD from structure B is 3.2 Å and that from structure C is 5.6 Å. This feature is consistent with the presence of helical-hairpin conformations only in the reduced ensemble of 20 representative conformations, which are most similar to PDB structure A (Figure 10A). With time averaging (τ = 489 ps), the RMSD distribution from structure A becomes broader, shifts to higher values, and largely overlaps with the RMSD distribution from structure B (Figure 11B), which is consistent with the presence of a continuity of structures between a helical hairpin and a partially open helical hairpin. It can also be seen that the RMSD from the kinked-helix structure C decreases, consistent with the presence of a kinked-helix structure in the reduced ensemble of 20 representative conformations (Figure 10D).

The interproton-distance restraints and restraint violations without time averaging and with time averaging for τ = 489 ps, n_{ave} = 1000 (calculated after the conversion of the coarse-grained structures to all-atom structures with cg2all[70,71]) are shown in Figure 12. The numerical values of the differences of the specific interproton distances from the upper distance boundaries from all runs are included in the `Suppdata.zip` archive of the Supporting Information. The ensemble-averaged right RMSDs from the upper interproton-distance boundaries, (ρ_u^+s; eq 17) and the percentages of satisfied distance restraints, calculated from all-atom structures for all runs, and from the PDB ensemble, are shown as bar plots in Figure 13A,B, respectively. The corresponding numerical values, and the values obtained from interproton distances estimated by ESCASA are collected in Table S1 of the Supporting Information.

The ensemble-averaged ρ_u^+s calculated from the all-atom structures obtained by backmapping with cg2all,[70,71] are 0.41 and 0.10 Å without and with time averaging (τ = 489 ps), respectively, and the numbers of violated restraints are 30 and 17, respectively; additionally the distance boundaries of 2 restraints are violated by more than 2 Å for the calculations run without time averaging. The ρ_u^+ values computed from the interproton distances estimated with ESCASA are 0.49 and 0.21 Å for the calculations without and with time averaging, respectively, and the numbers of distance-boundary violations are 50 (3 exceeding 2 Å) and 28 (none exceeding 2 Å), respectively. The ρ_u^+ calculated from the 2LWA PDB ensemble (a total of 60 structures) is 0.44 Å, with 41 violated distance restraints and 1 restraint violated beyond 2 Å. These violations are significantly larger than those obtained from our time-

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

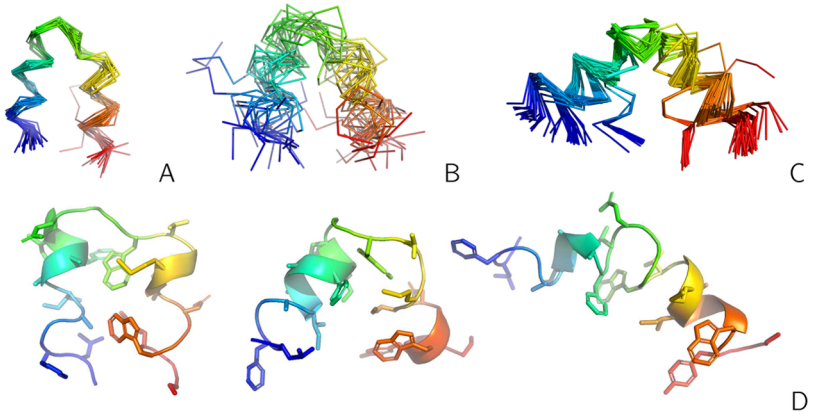


Figure 10. (A) Superposition of the C^{α} traces of the representative conformations of 20 families of 2LWA obtained in NMR-data-assisted UNRES/MREMD simulations without time averaging. (B) As in (A) but with time averaging (τ = 489 ps, n_{ave} = 1000). (C) Superposition of the C^{α} traces of all conformations of structures A, B, and C from the 2LWA PDB entry. (D) Three representative conformations of the ensemble shown in panel (B), which are closest to structure A, structure B, and structure C of the 2LWA PDB entry, respectively, shown in ribbon (for backbone) and stick (for side chains) representations. The chains are colored from blue to red from the N- to the C-terminus in all panels. The drawings were made with PyMOL.[60]

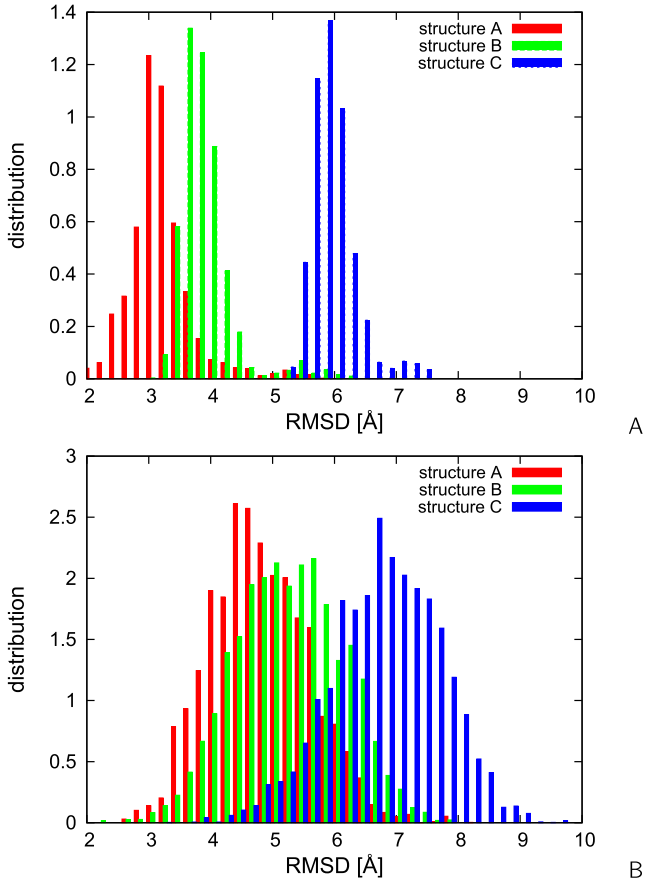


Figure 11. Distributions of C^{α}-RMSD of the structures of 2LWA from PDB structures A, B, and C for the conformational ensembles obtained in NMR-data-assisted MREMD simulations (A) without time averaging and (B) with time averaging. The plots were made with gnuplot.[79]

average-restraint data-assisted modeling with UNRES. Moreover, the ρ_u^+, the total number of violated distance restraints, and the number of restraints violated beyond 2 Å calculated from the subensemble of 20 structures obtained by clustering the ensemble obtained in this work with time averaging are 0.28 Å, 30, and 1, respectively, being still lower than those calculated from the conformations of the 2LWA PDB structure (Figure 13 and Table S1 of the Supporting Information).

Tests with Larger Proteins with Disordered Regions. In this part of our work, we carried out MREMD simulations of three proteins for which both NMR and X-ray structures are available: 2KW5 (3MER), 2KZN (3E0O), and 1PQX (2FFM) (cf. section "Systems Studied"). These proteins were part of the test set used in our previous work[53] to evaluate the performance of NMR-data-assisted UNRES simulations with interproton distances estimated using ESCASA.[40] For each of the three proteins, we carried out an MREMD simulation without time averaging and two simulations with time averaging, one with τ = 48.9 ps, n_{ave} = 500, and another one with τ = 489 ps, n_{ave} = 1000, respectively. In our previous work,[53] Hamiltonian replica exchange molecular dynamics (HREMD) simulations for these proteins were carried out. In HREMD, alteration of the energy function is another, apart from temperature, dimension of the replicas. Typically, the alteration is done by varying the weight of one or more of the energy components. Each of the temperatures is combined with each of the weights, making a 2D grid of replicas. In the HREMD simulations of ref 53, the restraint-penalty terms were assigned a weight varying from 0 to 1, constituting a total of 8 Hamiltonian replicas. The number of replica temperatures was 24 for 2KZN and 12 for 2KW5 and 2PQX. As for temperature replica exchange, the statistical weights of the conformations are obtained by postprocessing the resulting ensemble with WHAM. The final statistical weights of the conformations are calculated for the restraint-penalty weight of 1. Details can be found in ref 53. HREMD is more efficient (but also more resource-consuming) than MREMD and, therefore, running reference MREMD simulations without time averaging was necessary to assess the effect of time averaging on the results.

The ρ_u^+s from the upper distance boundaries, and the percentages of satisfied distance restraints are shown in Figure 13A,B, respectively. The values obtained by averaging over the representative conformations of 20 families obtained using the procedure described in section "Calculation Procedure" and those obtained by averaging over the structures of the respective PDB ensembles are also shown in Figure 13. The numerical values of the deviations from all restraints are collected in the respective machine-readable files of the `Suppinfo.zip`

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

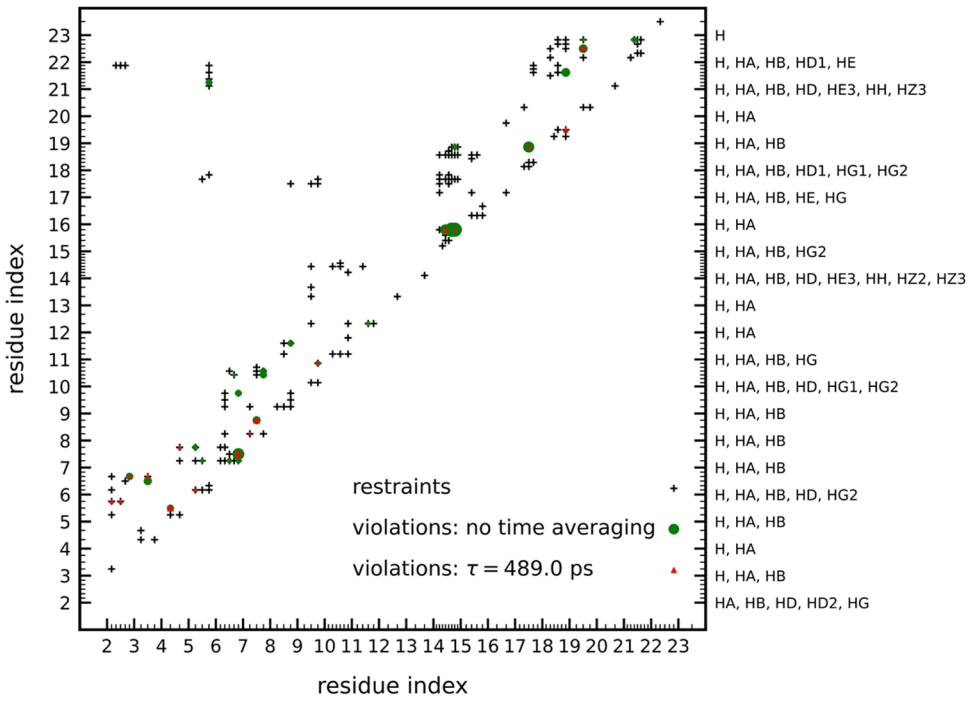


Figure 12. A diagram of NMR-determined interproton-distance restraints (black crosses) and violations of the upper boundaries of these restraints by ensemble-averaged interproton distances from NMR-assisted modeling of 2LWA with UNRES/MREMD (colored symbols). The size of the respective symbol is proportional to the deviation of the ensemble-averaged distance from the upper distance boundary. Green circles: simulations without time averaging. Red triangles: simulations with time averaging (τ = 489 ps, n_{ave} = 1000). The size of a symbol is proportional to the extent of upper-boundary violation. Each small tick corresponds to a proton of the residue with index below (ordinates) or to the left (abscissa) of it. The respective proton labels are shown on the right. The interproton distances of the consecutive conformations of the respective ensemble were calculated after conversion to all-atom structures.

archive. The ρ_u^+s and the numbers of violated restraints are collected in Table S1 of the Supporting Information, respectively. The diagrams of the interproton-distance restraints between pairs of residues and violations of the upper boundaries corresponding to the three calculation modes for the three proteins are shown in Figure 14A−C.

It can be seen from Figures 13 and 14 that the ρ_u^+s from the upper distance boundaries and the numbers of distance-boundary violations are smaller for the simulations with time averaging compared to those obtained in reference MREMD simulations. Only for 1PQX the ensemble obtained with τ = 48.9 ps is the ρ_u^+ higher than that from the reference simulation. However, the structure of 1PQX is well-defined by the restraints and, consequently, the ρ_u^+ and the numbers of violated restraints are small. The percentages of satisfied restraints (Figure 13B) show less regularity but they encompass only the satisfied restraints that are strictly below the upper boundaries, an entry even with an incremental violation being rejected. It can also be noted that increasing τ results in a greater number of ensemble-satisfied restraints. Even when the ensemble is reduced to 20 representative conformations the ρ_u^+ is lower and the percentage of satisfied distance restraints is greater than those from the PDB ensembles for 2KW5 and 2KZN (for the latter protein the differences are smaller). These results, as well as those obtained for 2LWA, suggest that, for multistate proteins and proteins with a significant amount of disordered regions, the advantage of extensive conformational search of the coarse-grained approach and restraint averaging overcomes the disadvantage of lower resolution inherent in a coarse-grained model.

The RMSDs of the top and best models from the X-ray structures and the GDT_TS values are shown as bar plots in Figure 15A−D and the numerical values are collected in Table S2 of the Supporting Information, respectively. The respective structures in the PDB format are included in the `Suppinfo.zip` archive of the Supporting Information. For this part of the analysis, the ensembles were dissected into 5 families, as in our previous work.[53] It can be seen that the RMSDs and GDT_TS are quite comparable for 1PQX, which again proves that the restraints define the structure of this protein well. Consequently, the resulting structure is less sensitive to the search method. On the other hand, unrestrained UNRES simulations result in a poor model of the structure of this protein, with C^{α}-RMSD = 10.5 Å and GDT_TS = 28.6 (ref 53), which demonstrates that the restraints are necessary to obtain a good model. For the two other proteins (2KW5 and 2KZN), which have disordered regions, restrained MREMD simulations without time averaging result in structures of the lowest quality, while MREMD simulations with time averaging result is structures of comparable or lower RMSD and comparable or higher GDT_TS than the more resource-consuming HREMD. This observation suggests that time averaging helps in searching the conformational space.

CONCLUSIONS

In this work, we have implemented the time-averaged restraints in NMR-data-assisted MD with the UNRES coarse grained model of polypeptide chains. Since the presence of time-dependent terms in the potential-energy function inevitably results in energy nonconservation, we developed a stable variant of the time-averaged MD algorithm, which involves updating the time averages every n_{ave} steps (n_{ave} usually ranging from 100 or 1000) with simple averages over this time window, and using the

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

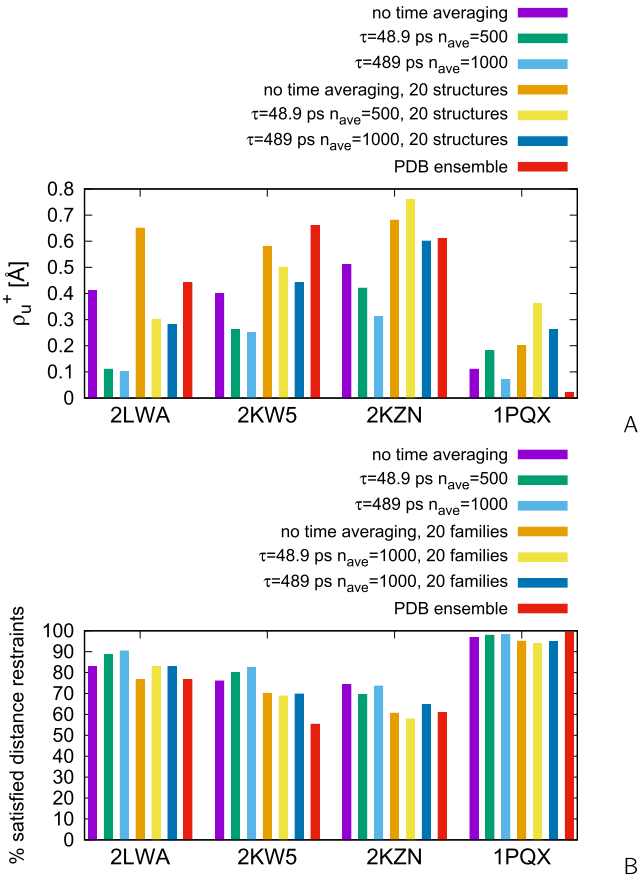


Figure 13. (A) Bar plot of the ρ_u^+ values of ensemble-averaged interproton distances from the right distance boundaries (eq 17). (B) Bar plot of the percentages of satisfied interproton-distance restraints for 2LWA, 1PQX, 2KW5, and 2KZN calculated from the results of MREMD simulations without time averaging and with time-averaged restraints, with τ = 48.9 ps, n_{ave} = 500 and τ = 489 ps, n_{ave} = 1000, respectively. The plots were made with gnuplot.[79]

average from the previous time windows and the current value of the conformation-dependent quantity under consideration in the other time steps (eqs 8−10). With this modification, the extended potential energy does not depend on trajectory history in each of the n_{ave}-step time windows, which provides total-energy conservation in each of these time periods and thus prevents the average kinetic temperature from deviating remarkably from the bath temperature in canonical runs, as opposed to updating the time averages every MD step (Table 2). We also found that scaling up the average restraint forces with the ratio of the memory-window-size (τ) to the MD time step (Δt) is necessary for the time-averaged restraints to affect the simulated structures (Figure 6). These two features are new with respect to the previously developed variants of the time-averaged-restraint methodology.[11−19] Compared to the methods based on replica-averaged restraints,[21−24] the time-averaged-restraint approach seems to provide more extensive averages because the memory-window length τ exceeds the MD time step many times (in our calculations from 1000 to 100,000 times), while the number of replicas is limited usually to several tens. However, because the NMR experiments result in both time- and ensemble-averaged observables,[9,10] the best approach should include both time- and replica-averaging. Such an approach is now being developed in our laboratory.

With the synthetic data of the small 2KW5(129−153) system, we demonstrated that time averaging results in broadening the distribution of conformations compared to distance-restrained simulations when the restraints are derived from a single reference structure (Figure 7). When the restraints are derived from two sufficiently distinct structures, only time averaging results in the presence of structures close to the first and those close to the second reference structure in simulations (Figures 8 and 9). Moreover, the ergodicity of canonical simulations is poor without time averaging when the input restraints originate from the averages over two or more conformations.

For 2LWA, which is a multistate protein, we obtained a more diverse conformational ensemble than that deposited in the PDB.[61] Instead of 3 distinct families of conformations present in the 2LWA PDB entry, one (structure A) characterized by a tightly packed helical hairpin, the second one (structure B) by an open helical hairpin, and the third one (structure C) by a kinked helix (Figure 3), our ensemble consists of a continuum of structures from tightly packed helical hairpins to kinked helices, with the gradually increasing angle between the two helices. Nevertheless the representative conformations of these boundary structural types constituting the PDB ensemble are present in our ensemble (Figure 10). Our ensemble results in a significantly better agreement of the calculated and experimental interproton distances than the PDB ensemble (Figure 13 and Table S1 of the Supporting Information). It should be noted that, in our coarse-grained simulations, the proton positions were estimated with ESCASA and only the final ensembles were converted to the all-atom representation from which the final interproton distances were calculated. Therefore, the agreement with the experimental interproton distances could presumably be improved if all-atom refinement subject to NMR-distance-restraints in the time-average mode was carried out.

Owing to time- and size-scale extension of coarse-grained simulations with respect to all-atom simulations, we could try modeling, with time-averaged restraints, the structures of larger proteins. We selected three proteins, one of which has a well-defined structure (1PQX), while the two other ones (2KW5 and 2KZN) have disordered regions. For 1PQX, the NMR ensemble obtained with routine data processing conforms with the experimental data better than the ensembles from our calculations. However, for the two other proteins, our calculations give better agreement with the experimental NMR restraints. Thus, when the restraints define the structure well, all-atom protein-structure modeling is advantageous over coarse-grained modeling, because the interproton distances are only estimated in the coarse-grained approach. This conclusion has already been drawn in our earlier study.[53] However, the coarse-grained approach with time averaging seems more robust for proteins with disordered regions (Figure 13). As the results of the calculations for 2KW5 and 2KZN also suggest, introducing time averaging improves the ergodicity of simulations, because the structures obtained from time-average simulations for these two proteins had similar or higher GDT_TS and a lower or similar RMSD with respect to the reference X-ray structures (Figure 15) than those obtained in our earlier work using the more resource-intensive HREMD method.[53] Therefore, coarse-grained data-assisted modeling with time-averaged restraints can be a method of choice for larger proteins with disordered regions, for which all-atom modeling in the time-average mode could be infeasible due to insufficient search of the conformational space.

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

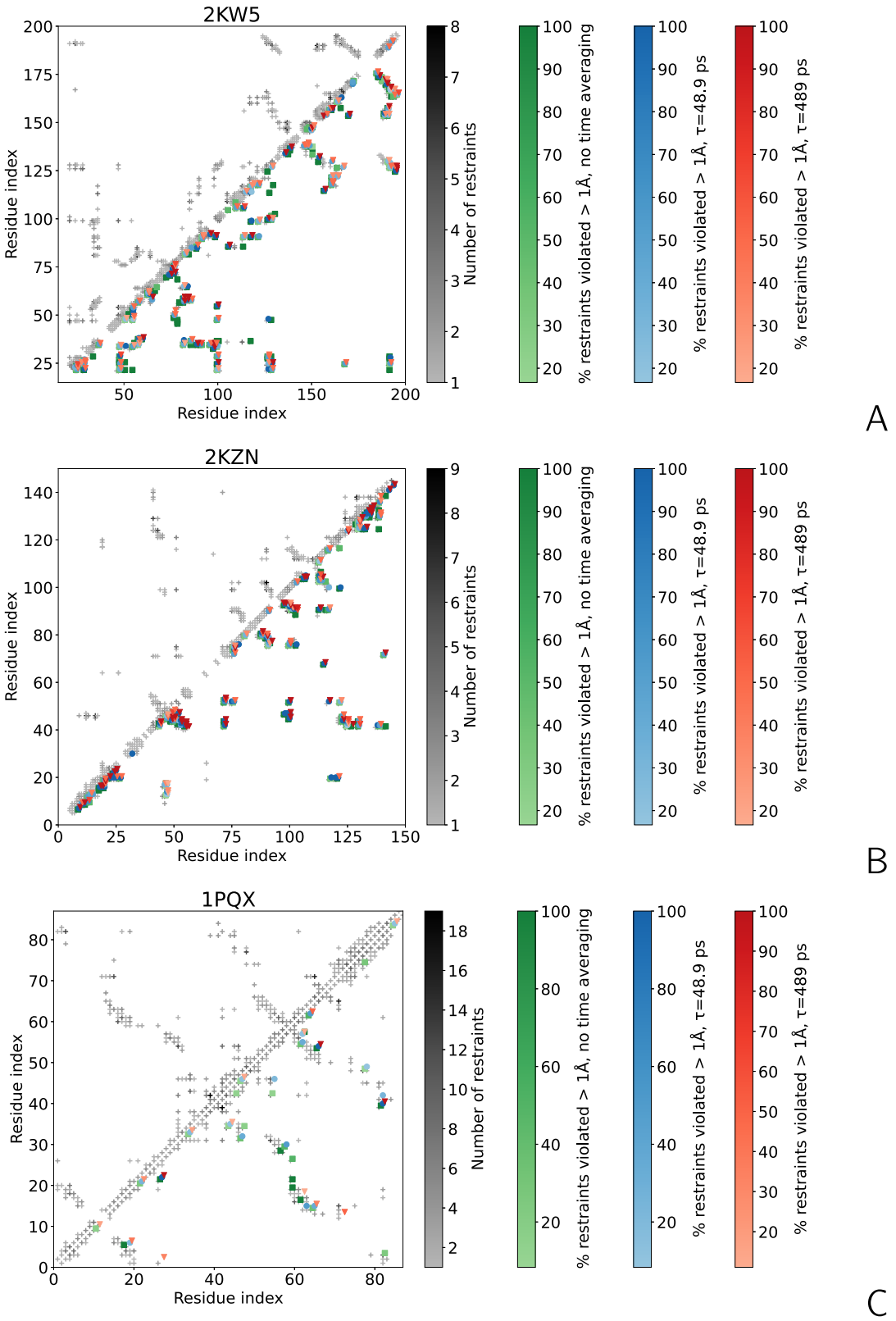


Figure 14. Diagrams of NMR-determined interproton-distance restraints and ensemble-averaged distance-restraint violations of the conformational ensembles obtained from MREMD simulations for (A) 2KW5, (B) 2KZN, and (C) 1PQX. The interproton-distance restraints pertaining to pairs of residues are shown as crosses in the upper-diagonal part, with the degree of gray proportional to the number of interproton-distance restraints pertaining to a pair. Restraint violations are shown in the lower-diagonal part as green circles, for the simulations without time averaging, blue squares, for the simulations with time averaging, τ = 48.9 ps, n_{ave} = 500, and red triangles for the simulations with time averaging, τ = 489 ps, n_{ave} = 1000, the color saturation proportional to the percentage of violated restraints for a given pair. Only the violations no less than 1 Å are shown. The crosses in the lower-diagonal part indicate the pairs of residues where no violations were observed.

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

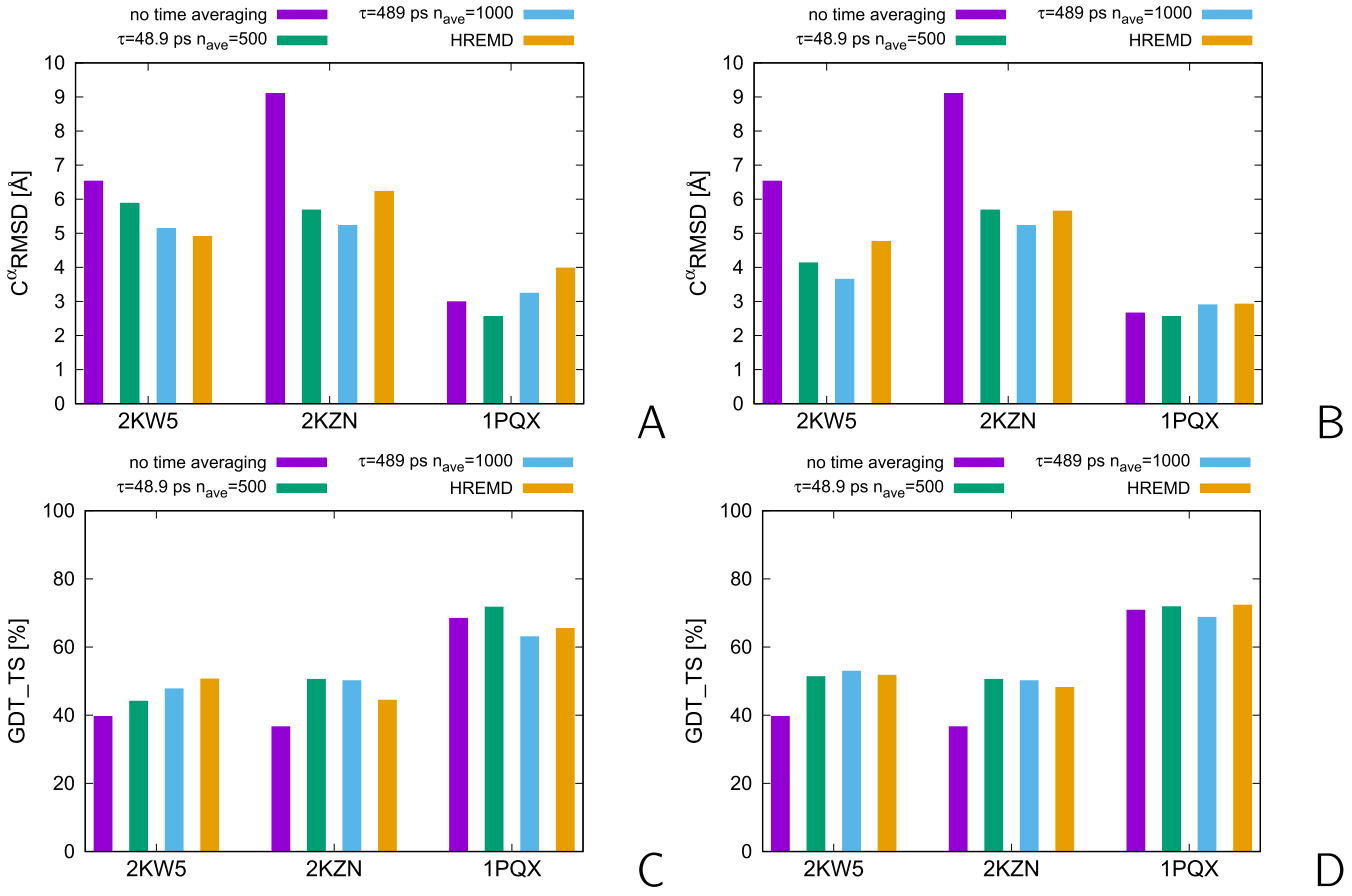


Figure 15. (A) Bar plot of C^{α}-RMSDs from the reference X-ray structures for the top models of 1PQX, 2KW5, and 2KZN obtained in MREMD NMR-data-assisted simulations without time averaging and with time averaging, with τ = 48.9 ps, n_{ave} = 500, and τ = 489 ps, n_{ave} = 1000, respectively, and those obtained in HREMD simulations of our previous work.[53] (B) Bar plot of the C^{α}-RMSDs for the best models. (C) Bar plot of the GDT_TS of the first models (with X-ray structures as reference structures). (D) Bar plot of the GDT_TS for the best models. The plots were made with gnuplot.[79]

Although the ensembles found by our NMR-data-assisted coarse-grained protein-structure-modeling approach are, except for 1PQX, in a better agreement with the experimental data than the ensembles deposited in the PDB, they contain much more conformations. Our attempts at cutting down the number of conformations to a typical NMR-determined protein-ensemble size from the PDB, by dissecting the ensemble into 20 families and selecting the best-fitting representative of each cluster resulted in ensembles still better fitting the NMR data than the corresponding PDB ensembles, except for 1PQX. Nevertheless, ensemble reweighting or limited time-average NMR-data-assisted modeling at the all-atom level are likely to result in a still better fitting. The work on this problem, as well the work on introducing explicit ϕ and ψ backbone-dihedral-angle restraints via estimating these angles from the coarse-grained geometry, is now being carried out in our laboratory.

■ ASSOCIATED CONTENT

Data Availability Statement

The UNRES software with time-average-modeling capacity is available at https://unres.pl/downloads under the GPL v3 license. The interproton-distance and angular restraints, the PDB files of the simulated structures, and the violations of the upper distance boundaries are in the machine-readable format in the `Suppdata.zip` archive, which is a part of the Supporting Information. The values of the ρ_u^+ and the numbers of upper-boundary violations are summarized in Table S1 and the C^{α}-RMSD and GDT_TS values from the reference X-ray structures for the models of 1PQX, 2KW5, and 2KZN are in Table S2 of the Supporting Information.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c01504.

Section "Glossary of the machine-readable files". Short description of the content of the `Suppdata.zip` containing machine-readable files. Table S1. ρ_u^+s from the upper distance boundaries, the total numbers of violated restraints, and the numbers of restraints violated by more than 2 Å for the 2LWA, 1PQX, 2KW5, and 2KZN proteins. Table S2. C^{α}-RMSD and GDT_TS values from the respective X-ray structures of the top and best models of 2KW5, 2KZN, and 1PQX obtained in NMR-data-assisted UNRES/MREMD simulations without time averaging and with time averaging (PDF)

Machine-readable files with distance and angular restraints, models of the proteins under study obtained from simulations, and NMR-restraint violations (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Adam Liwo − *Faculty of Chemistry, University of Gdańsk, Fahrenheit Union of Universities, 80-308 Gdańsk, Poland;* orcid.org/0000-0001-6942-2226; Phone: +48 58

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

5235124; Email: adam.liwo@ug.edu.pl; Fax: +48 58 5235012

Authors

Nguyen Truong Co − *Faculty of Chemistry, University of Gdańsk, Fahrenheit Union of Universities, 80-308 Gdańsk, Poland;* orcid.org/0000-0001-5642-3641

Cezary Czaplewski − *Faculty of Chemistry, University of Gdańsk, Fahrenheit Union of Universities, 80-308 Gdańsk, Poland;* orcid.org/0000-0002-0294-3403

Emilia A. Lubecka − *Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Fahrenheit Union of Universities in Gdańsk, 80-233 Gdańsk, Poland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c01504

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Science Centre of Poland under grant UMO-2021/40/Q/ST4/00035. Computational resources were provided by (a) the Centre of Informatics - Tricity Academic Supercomputer & Network (CI TASK) in Gdańsk, (b) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) the University of Warsaw under grant No. GA71-23 and (d) our 796-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

■ REFERENCES

(1) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964−972.
(2) Boehr, D. D.; Nussinov, R.; Wright, P. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(3) Bertelsen, E. B.; Chang, L.; Gestwick, J. E.; Zuiderweg, E. R. P. Solution Conformation of Wild-Type *E. coli* Hsp70 (DnaK) Chaperone Complexed with ADP and Substrate. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 8471−8476.
(4) Kityk, R.; Kopp, J.; Sinning, I.; Mayer, M. P. Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. *Mol. Cell* **2012**, *48*, 863−874.
(5) van der Lee, R.; Buljan, M.; Lang, B.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589−6631.
(6) Uversky, V. N. *Intrinsically Disordered Proteins*; Springer, 2016; Vol. *4*, e1135015.
(7) Dunker, A. K.; Lawson, J.; Brown, C. J.; et al. Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.
(8) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(9) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132*, 8407−8418.
(10) Konrat, R. NMR Contributions to Structural Dynamics Studies of Intrinsically Disordered Proteins. *J. Magn. Reson.* **2014**, *241*, 74−85.
(11) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time-Dependent Distance Restraints in Molecular Dynamics Simulations. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(12) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time-Averaged Nuclear Overhauser Effect Distance Restraints Applied to Tendamistat. *J. Mol. Biol.* **1990**, *214*, 223−235.
(13) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. Structure Refinement Using Time-Averaged J-Coupling Constant Restraints. *J. Biomol. NMR* **1993**, *3*, 55−66.
(14) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. Time- and Ensemble-Averaged Direct NOE Restraints. *J. Biomol. NMR* **1994**, *4*, 143−149.
(15) Nanzer, A. P.; van Gunsteren, W. F.; Torda, A. E. Parametrisation of Time-Averaged Distance Restraints in MD Simulations. *J. Biomol. NMR* **1995**, *6*, 313−320.
(16) Scott, W. R. P.; Mark, A. E.; van Gunsteren, W. F. On Using Time-Averaging Restraints in Molecular Dynamics Simulations. *J. Biomol. NMR* **1998**, *12*, 501−508.
(17) Dolenc, J.; Missimer, J. H.; Steinmetz, M. O.; van Gunsteren, W. F. Methods of NMR Structure Refinement: Molecular Dynamics Simulations Improve the Agreement with Measured NMR Data of a C-Terminal Peptide of GCN4-p1. *J. Biomol. NMR* **2010**, *47*, 221−235.
(18) Smith, L. J.; van Gunsteren, W. F.; Hansen, N. On the Use of Time-Averaging Restraints when Deriving Biomolecular Structure from ^{3}J-Coupling Values Obtained from NMR Experiments. *J. Biomol. NMR* **2016**, *66*, 69−83.
(19) Tzvetkova, P.; Sternberg, U.; Gloge, T.; Navarro-Vázquez, A.; Luy, B. Configuration Determination by Residual Dipolar Couplings: Accessing the Full Conformational Space by Molecular Dynamics with Tensorial Constraints. *Chem. Sci.* **2019**, *10*, 8774−8791.
(20) Harish, B.; Swapna, G. V. T.; Kornhaber, G.; Montelione, G. T.; Carrey, J. Multiple Helical Conformations of the Helix-Turn-Helix Region Revealed by NOE-Restrained MD Simulations of Tryptophan Aporepressor, TrpR. *Proteins:Struct., Funct., Bioinf.* **2017**, *85*, 731−740.
(21) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular Dynamics Simulations with Replica-Averaged Structural Restraints Generate Structural Ensembles According to the Maximum Entropy Principle. *J. Chem. Phys.* **2013**, *138*, No. 094112.
(22) Olson, M. A.; Lee, M. S. Application of Replica Exchange Umbrella Sampling to Protein Structure Refinement of Nontemplate Models. *J. Comput. Chem.* **2013**, *34*, 1785−1793.
(23) Roux, B.; Weare, J. On the Statistical Equivalence of Restrained-Ensemble Simulations with Maximum Entropy Method. *J. Chem. Phys.* **2013**, *138*, No. 084107.
(24) Hummer, G.; Köfinger, J. Bayesian Ensemble Refinement by Replica Simulations and Reweighting. *J. Chem. Phys.* **2015**, *143*, No. 243150.
(25) Nikiforovich, G. V.; Vesterman, B.; Betins, J.; Podins, L. The Space Structure of a Conformationally Labile Oligopeptide in Solution - Angiotensin. *J. Biomol. Struct. Dyn.* **1987**, *4*, 1119−1135.
(26) Groth, M.; Malicka, J.; Czaplewski, C.; Ołdziej, S.; Łankiewicz, L.; Wiczk, W.; Liwo, A. Maximum Entropy Approach to the Determination of Solution Conformation of Flexible Polypeptides by Global Conformational Analysis and NMR Spectroscopy - Application to DNS1-c-[D-A2bu2,Trp4,Leu5]enkephalin and DNS1-c-[D-A2bu2,Trp4,D-Leu5]enkephalin. *J. Biomol. NMR* **1999**, *15*, 315−330.
(27) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian Inference Method for Heterogeneous Systems. *Sci. Adv.* **2016**, *2*, No. e1501177.
(28) Schwieters, C. D.; Bermejo, G. A.; Clore, G. M. Xplor-NIH for Molecular Structure Determination from NMR and Other Data Sources. *Protein Sci.* **2018**, *27*, 26−40.
(29) Selegato, D. M.; Bracco, C.; Giannelli, C.; Parigi, G.; Luchinat, C.; Sgheri, L.; Ravera, E. Comparison of Different Reweighting Approaches for the Calculation of Conformational Variability of Macromolecules from Molecular Simulations. *ChemPhysChem* **2021**, *22*, 127−138.
(30) Tozzini, V. Minimalist Models of Proteins: A Comparative Analysis. *Q. Rev. Biophys.* **2010**, *43*, 333−337.
(31) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898−7936.
(32) Noid, W. G. Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models. *J. Phys. Chem. B* **2023**, *127*, 4174−4207.

(33) Borges-Araújo, L.; Patmanidis, I.; Singh, A. P.; Santos, L. H. S.; Sieradzan, A. K.; Vanni, S.; Czaplewski, C.; Pantano, S.; Shinoda, W.; Monticelli, L.; Liwo, A.; Marrink, S. J.; Souza, P. C. T. Pragmatic Coarse-Graining of Proteins: Models and Applications. *J. Chem. Theory Comput.* **2023**, *19*, 7112−7135.
(34) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α-Helical Systems. *J. Phys. Chem. B* **2005**, *109*, 13798−13810.
(35) Sieradzan, A. K.; Sans-Duñó, J.; Lubecka, E. A.; Czaplewski, C.; Lipska, A. G.; Leszczynski, H.; Ocetkiewicz, K. M.; Proficz, J.; Czarnul, P.; Krawczyk, H.; Liwo, A. Optimization of Parallel Implementation of UNRES Package for Coarse-Grained Simulations to Treat Large Proteins. *J. Comput. Chem.* **2023**, *44*, 602−625.
(36) Liwo, A.; Baranowski, M.; Czaplewski, C.; et al. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20*, No. 2306.
(37) Sieradzan, A. K.; Czaplewski, C.; Krupa, P.; Mozolewska, M. A.; Karczyńska, A. S.; Lipska, A. G.; Lubecka, E. A.; Gołaś, E.; Wirecki, T.; Makowski, M.; Ołdziej, S.; Liwo, A. *Protein Folding: Methods and Protocols*; Muñoz, V., Ed.; Springer US: New York, NY, 2022; pp 399−416.
(38) Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lubecka, E. A.; Lipska, A. G.; Golon, L.; Karczyńska, A.; Krupa, P.; Mozolewska, M. A.; Makowski, M.; Ganzynkowicz, R.; Giełdoń, A.; Maciejczyk, M. Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly. In *Progress in Molecular Biology and Translational Science*; Strodel, B.; Barz, B., Eds.; Academic Press, 2020; Vol. *170*, pp 73−122.
(39) Antoniak, A.; Biskupek, I.; Bojarski, K. K.; et al. Modeling Protein Structures with the Coarse-Grained UNRES Force Field in the CASP14 Experiment. *J. Mol. Graphics Modell.* **2021**, *108*, No. 108008.
(40) Lubecka, E. A.; Liwo, A. ESCASA: Analytical Estimation of Atomic Coordinates from Coarse-Grained Geometry for Nuclear-Magnetic-Resonance-Assisted Protein Structure Modeling. I. Backbone and H^{β} Protons. *J. Comput. Chem.* **2021**, *42*, 1579−1589.
(41) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(42) Sieradzan, A. K.; Makowski, M.; Augustynowicz, A.; Liwo, A. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. I. Backbone Potentials of Coarse-Grained Polypeptide Chains. *J. Chem. Phys.* **2017**, *146*, No. 124106.
(43) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. Cumulant-Based Expressions for the Multibody Terms for the Correlation between Local and Electrostatic Interactions in the United-Residue Force Field. *J. Chem. Phys.* **2001**, *115*, 2323−2347.
(44) Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Żmudzińska, W.; Hałabis, A.; Ołdziej, S. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-Field Calibration and Validation. *J. Chem. Phys.* **2019**, *150*, No. 155104.
(45) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. I. Lagrange Equations of Motion and Tests of Numerical Stability in the Microcanonical Mode. *J. Phys. Chem. B* **2005**, *109*, 13785−13797.
(46) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. Application of Multiplexing Replica Exchange Molecular Dynamics Method to the UNRES Force Field: Tests with α and $\alpha + \beta$ Proteins. *J. Chem. Theory Comput.* **2009**, *5*, 627−640.
(47) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S.; Kaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. Atomistic Protein Folding Simulations on the Submillisecond Time-scale Using Worldwide Distributed Computing. *Biopolymers* **2003**, *68*, 91−109.
(48) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(49) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. Modification and Optimization of the United-Residue (UNRES) Potential Energy Function for Canonical Simulations. I. Temperature Dependence of the Effective Energy Function and Tests of the Optimization Method with Single Training Proteins. *J. Phys. Chem. B* **2007**, *111*, 260−285.
(50) Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kleinerman, D. S.; Blood, P.; Scheraga, H. A. Implementation of Molecular Dynamics and its Extensions With the Coarse-Grained UNRES Force Field on Massively Parallel Systems; Towards Millisecond-Scale Simulations of Protein Structure, Dynamics, and Thermodynamics. *J. Chem. Theory Comput.* **2010**, *6*, 583−595.
(51) Ocetkiewicz, K. M.; Czaplewski, C.; Krawczyk, H.; Lipska, A. G.; Liwo, A.; Proficz, J.; Sieradzan, A. K.; Czarnul, P. UNRES-GPU for Physics-Based Coarse-Grained Simulations of Protein Systems at Biological Time- and Size-Scales. *Bioinformatics* **2023**, *39*, No. btad391.
(52) Ocetkiewicz, K. M.; Czaplewski, C.; Krawczyk, H.; Lipska, A. G.; Liwo, A.; Proficz, J.; Sieradzan, A. K.; Czarnul, P. Multi-GPU UNRES for Scalable Coarse-Grained Simulations of Very Large Protein Systems. *Comput. Phys. Commun.* **2024**, *298*, No. 109112.
(53) Lubecka, E.; Liwo, A. A Coarse-grained Approach to NMR-Data-Assisted Modeling of Protein Structures. *J. Comput. Chem.* **2022**, *43*, 2047−2059.
(54) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. Low-Energy Structures of Two Dipeptides and Their Relationship to Bend Conformations. *Macromolecules* **1974**, *7*, 797−806.
(55) Krupa, P.; Mozolewska, M. A.; Wiśniewska, M.; et al. Performance of Protein-Structure Predictions with the Physics-Based UNRES Force Field in CASP11. *Bioinformatics* **2016**, *32*, 3270−3278.
(56) Lubecka, E. A.; Karczyńska, A. S.; Lipska, A. G.; et al. Evaluation of the Scale-Consistent UNRES Force Field in Template-Free Prediction of Protein Structures in the CASP13 Experiment. *J. Mol. Graphics Modell.* **2019**, *92*, 154−166.
(57) Ślusarz, R.; Lubecka, E.; Czaplewski, C.; Liwo, A. Improvements and New Functionalities of UNRES Server for Coarse-Grained Modeling of Protein Structure, Dynamics, and Interactions. *Front. Biomol. Sci.* **2022**, *9*, No. 1071428.
(58) Pearlman, D. A.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, T., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(59) Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y.; Lee, H. W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Determination of Solution Structures of Proteins up to 40 kDa Using CS-Rosetta with Sparse NMR Data from Deuterated Samples. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10873−10878.
(60) Schrödinger, L. L. C. The PyMOL molecular graphics system. 2010.
(61) Lorieau, J. L.; Louis, J. M.; Schwieters, C. D.; Bax, A. pH-Triggered, Activated-State Conformations of the Influenza Hemagglutinin Fusion Peptide Revealed by NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 19994−19999.
(62) Everett, J. K.; Tejero, R.; Murthy, S. B. K.; et al. A Community Resource of Experimental Data for NMR/X-Ray Crystal Structure Pairs. *Protein Sci.* **2016**, *25*, 30−45.
(63) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(64) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493

Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637−649.
(65) Lee, J.; Liwo, A.; Scheraga, H. A. Energy-Based *De Novo* Protein Folding by Conformational Space Annealing and an Off-Lattice United-Residue Force Field: Application to the 10−55 Fragment of Staphylococcal Protein A and to *Apo*-Calbindin D9K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2025−2030.
(66) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. A United-Residue Force Field for Off-Lattice Protein-Structure Simulations. II: Parameterization of Local Interactions and Determination of the Weights of Energy Terms by Z-score Optimization. *J. Comput. Chem.* **1997**, *18*, 874−887.
(67) Trebst, S.; Troyer, M.; Hansmann, U. H. E. Optimized Parallel Tempering Simulations of Proteins. *J. Chem. Phys.* **2006**, *124*, No. 174903.
(68) Murtagh, F.; Heck, A. *Multivariate Data Analysis*; Kluwer Academic Publishers, 1987.
(69) Protein Structure Prediction Center. 2024 https://predictioncenter.org/. (accessed on November 5, 2024).
(70) Heo, L.; Feig, M. One Bead per Residue Can Describe All-Atom Protein Structures 2023 https://github.com/huhlim/cg2all. (accessed on November 5, 2024).
(71) Heo, L.; Feig, M. One Bead Per Residue Can Describe All-Atom Protein Structures. *Structure* **2024**, *32*, 97−111.
(72) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198−210.
(73) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.
(74) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3*, 156−169.
(75) Huang, H.; Simmerling, C. Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. *J. Chem. Theory Comput.* **2018**, *14*, 5797−5814.
(76) Zemla, A.; Venclovas, C.; Fidelis, K.; Rost, B. A Modified Definition of SOV, a Segment-Based Measure for Protein Secondary Structure Prediction Assessment. *Proteins Struct. Funct. Genet.* **1999**, *34*, 220−223.
(77) Zemla, A. LGA: a Method for Finding 3D Similarities in Protein Structures. *Nucleic Acids Res.* **2003**, *31*, 3370−3374.
(78) Zhang, Y.; Skolnick, J. Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 702−710.
(79) Williams, T.; Kelley, C.et al. Gnuplot 4.6: An Interactive Plotting Program 2013 http://gnuplot.sourceforge.net/. (accessed on November 5, 2024).
(80) Kelley, D.; Galbraith, P. GRI Version 2.12.23 2023 http://gri.sourceforge.net/. (accessed on November 5, 2024).

https://doi.org/10.1021/acs.jctc.4c01504
J. Chem. Theory Comput. 2025, 21, 1476−1493