JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Accelerating the Conformational Sampling of Intrinsically Disordered Proteins

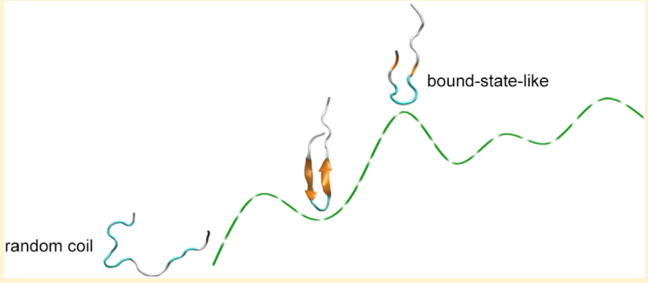
Trang Nhu Do,[†] Wing-Yiu Choy,[‡] and Mikko Karttunen*,[†]

[†]Department of Chemistry and Waterloo Institute for Nanotechnology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

[‡]Department of Biochemistry, University of Western Ontario, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

Supporting Information

ABSTRACT: Intrinsically disordered proteins (IDPs) are a class of proteins lacking a well-defined secondary structure. Instead, they are able to attain multiple conformations, bind to multiple targets, and respond to changes in their surroundings. Functionally, IDPs have been associated with molecular recognition, cell regulation, and signal transduction. The dynamic conformational ensemble of IDPs is highly environmental and binding partner dependent, rendering the characterization of IDPs extremely challenging. Here, we compare the sampling efficiencies of conventional molecular dynamics (MD), well-tempered metadynamics (WT-META), and bias-exchange metadynamics (BE-META). The total simulation time was over 10 μs, and a 20-mer peptide derived from the Neh2 domain of the Nuclear factor erythroid 2-related factor 2 (Nrf2) protein was simulated. BE-META, with a neutral replica and seven biased replicas employing a set of seven relevant collective variables (CVs), provided the most reliable and efficient sampling. Finally, we propose a free-energy reconstruction method based on the probability distribution of the secondary structure contents. This postprocessing analysis confirms the presence of not only the β-hairpin conformation of the free Neh2 peptide but also its rare bound-state-like conformation, both of that have been experimentally observed. In addition, our simulations also predict other possible conformations to be verified with future experiments.



1. INTRODUCTION

Intrinsically disordered proteins (IDPs) do not have a well-defined secondary or tertiary structure in their native unbound states under physiological conditions.[1−4] Instead, IDPs can adopt a wide spectrum of configurations from fully to partially disordered in a form of dynamic ensemble of states. This structural plasticity allows IDPs to undergo effective conformational transitions and hence acquire excellent complementarities with a wide range of binding surfaces of their targets:[5,6] IDPs can fold from a disordered to somewhat ordered states upon interacting with their binding partners.[6−11] Despite the lack of well-defined structures, IDPs are essential for numerous biological functions, especially those related to molecular recognition, cell regulation, and signal transduction,[3,5,12,13] and they are involved in cancers, aging, cardiovascular diseases, amyloidosis, and neurodegenerative diseases.[14−16] It has been shown that up to 35−40% of proteins are either disordered or contain significant disordered regions[17,18] and IDPs are specifically conserved by evolution.[6,19] Although the importance and abundance of IDPs is now well-established, their behavior and mechanisms involved in folding and binding remain underexplained.[8−10,19−21]

One of the fundamental obstacles in characterization is that when not binding to a target, IDPs exist as vast dynamic ensembles of conformations that, importantly, cannot be easily determined from an ensemble average.[6,22] This challenges traditional biophysical approaches including nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. Other techniques such as fluorescence resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) are able to provide valuable information for IDPs but still suffer from inaccuracy.[23,24] Combining experiments with computer simulations has proven to be a fruitful approach.[25,26]

Computer simulations of IDPs have two main challenges: (i) the quality of force fields and (ii) computational cost. As for the former, some force fields have been found to favor either α-helix or β-sheet structures.[27−31] A lot of effort has, however, been invested to test and improve them.[29,32−39] For the latter, the high computational cost arises from the complexity of the conformational space of IDPs and difficulty of efficiently sampling the dynamic ensemble of states. Instead of unbiased molecular dynamics (MD) simulations, accelerated techniques have led to dramatic improvements in sampling the conformational space of proteins. These methods include path-based approaches (transition-path sampling,[40−43] path-optimization techniques,[44,45] etc.), equilibrium-distribution approaches (umbrella sampling,[46] replica exchange MD,[47] etc.), and other non-

Received: June 3, 2014
Published: October 10, 2014

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

equilibrium approaches (steered MD,[48,49] metadynamics,[50] etc.). Of these, path-based approaches require a priori knowledge of the end states and/or the associated transition path. Therefore, they are less applicable for IDPs. Replica exchange MD (REMD) approaches can be powerful in sampling the conformational changes of IDPs. However, the number of replicas, and thus the demanded computational resources, scale with the system size, making REMD less practical for large IDPs and IDP complexes. Methods based on a single reaction coordinate or collective variable (CV) such as umbrella sampling and steered MD require the choice of a proper CV to bias the simulations. A single fully descriptive CV does not exist for such a complex state transition of IDPs. Metadynamics technique appears as a suitable enhanced sampling method to explore the conformational space of IDPs. One of its advantages is that knowledge of the end states is not required. Moreover, metadynamics allows biasing on more than one CVs. A recent metadynamics variant, the so-called bias-exchange metadynamics,[51] combines the standard metadynamics method[50] and the replica exchange approach[47] into a single method allowing a powerful exploration of the free energy landscape in a virtually unlimited number of collective variables. The bias-exchange metadynamics approach with careful choices of a set of biasing CVs have been successfully applied to study the conformational space of proteins including IDPs.[52−55]

The IDP we used was the Neh2 domain of the Nuclear factor erythroid 2-related factor 2 (Nrf2) protein. Nrf2 is an important transcription factor responsible for primary cellular defense against the cytotoxicity caused by oxidative stress.[56,57] Neh2 is the N-terminal domain of Nrf2, and it is intrinsically disordered.[58] Available X-ray crystallography data (PDB: 2FLU) show that the Neh2-derived peptide (residues 75−83) containing the higher affinity "ETGE" motif forms a β-hairpin conformation when in complex with the Kelch domain of the Kelch-like ECH-associated protein 1 (Keap1).[59] NMR data of the full length Neh2 protein suggest that the region consisting of residues LEU74-PRO85 likely forms a more extended β-hairpin structure when Neh2 is in its free state.[58]

We performed simulations of a 20-mer human Neh2 domain (^{71}FAQLQLDEETGEFLPIQPAQ^{90}). The peptide contains the "ETGE" motif essential for Kelch binding and has a sufficient length for any secondary structure formation. Our work contains an extensive comparison of the conformational sampling by a 3-μs conventional unbiased MD simulation, a 3-μs well-tempered metadynamics[60] biasing on two CVs, and a 460 ns bias-exchange metadynamics[51] with eight replicas including seven biased and a neutral replica. The results show that with a comparable simulation time and computational effort, bias-exchange metadynamics with a larger set of relevant biasing CVs is the most efficient and sufficient technique of these three methods. Our results also suggest that the highly complex conformational space of the disordered Neh2 peptide requires at least seven CVs. Moreover, we propose a postprocessing method for free energy reconstruction. This method is based on the description of the secondary structures instead of the biasing CVs as in conventional free-energy reconstruction. The analysis is motivated by the observation that all seven employed CVs in our bias-exchange metadynamics are critically important for the exploration of the free energy profile but none of them as a single CV is adequate for the reconstruction of the free energy profile. Our analysis results in a secondary-structure-based free energy reconstruction, in which similar secondary structures are grouped together to ensure an accurate probability distribution of conformations. The conformational ensemble obtained for the Neh2 peptide from our simulations is validated against the available X-ray crystallography and NMR data.[58,59] Our simulations reproduce the X-ray conformation of Neh2 in its bound state with Kelch[59] and provide good agreement with NMR data of Neh2 in its free state.[58] We provide further information on the stability of the Neh2 bound-state-like conformation when Neh2 is not in complex with Kelch. We also suggest other highly probable folded conformations of Neh2 that have not yet been observed experimentally or computationally.

2. MATERIALS AND METHODS

2.1. Folding Simulations with Metadynamics. Metadynamics is an accelerated-sampling simulation technique. It enhances the probability of observing state transitions hence provides a better sampling of the conformational space and can be used for free-energy calculations. Metadynamics involves adding a history-dependent biasing potential to accelerate the occurrence of rare events and hence, unlikely states have enhanced chances to be explored during the simulations.[61−63] Eventually, the biasing potential compensates the underlying free energy surface that can then be reconstructed from the deposited biasing potential.[64] In this work, we employ two variants of metadynamics, well-tempered metadynamics (WT-META)[60] and bias-exchange metadynamics (BE-META).[51]

2.1.1. Well-Tempered Metadynamics (WT-META). Metadynamics is based on the dynamics performed by a few collective variables of the atomic coordinates $s(\mathbf{r})$.[50] In a metadynamics simulation, the dynamics of the system is driven not only by the potential energy of the system but also by a biasing history-dependent potential constructed as a sum of Gaussian potentials centered along the trajectory of the collective variables $s(\mathbf{r})$[50]

$$V_G(s(\mathbf{r}), t) = w \sum_{\substack{t'=\tau_G,2\tau_G,\ldots \\ t'<t}} \exp\left(-\frac{(s(\mathbf{r}) - s(t'))^2}{2\sigma_s^2}\right) \quad (1)$$

where w and σ_s are the height and the width of the Gaussian potentials, and τ_G is the rate of their deposition. In standard metadynamics,[50] the Gaussian height w is kept constant during the simulation. Well-tempered metadynamics (WT-META) is a variant of standard metadynamics that adds an adaptive bias along the trajectory of the collective variables by varying the Gaussian height,[60]

$$w(s(t)) = w_0 \frac{\Delta T}{\Delta T + w_0 N(s, t)} \tau_G \quad (2)$$

where w_0 is the initial Gaussian height, $N(s, t)$ denotes the number of times the system visits state s up to time t, and ΔT is a temperature parameter that can be tuned to increase the probability of barrier crossing. The free-energy exploration is automatically limited to an energy range of the order of the so-called enhanced temperature $T + \Delta T$,[60] meaning that with the WT-META scheme the computational effort is naturally focused on the physically relevant regions of the collective variables.

2.1.2. Bias-Exchange Metadynamics (BE-META). Bias-exchange metadynamics (BE-META) combines the standard metadynamics[50] with replica-exchange.[47] This allows an exploration of the free energy landscape with virtually unlimited number of variables.[51] In the conventional replica-exchange[47] and the parallel-tempering metadynamics,[65] different replicas at different temperatures are allowed to periodically exchange conformation with an exchange probability. In BE-META,

different replicas are simulated at the same temperature but biased by different history-dependent potentials acting on different collective variables. Analogously to standard metadynamics, the biasing potential $V_G^i(\mathbf{r}, t)$ acting on the collective variable $s^i(\mathbf{r})$ sampled by replica i is given by

$$V_G^i(\mathbf{r}, t) = V_G(s^i(\mathbf{r}), t) = w \sum_{\substack{t'=\tau_G,2\tau_G,\ldots \\ t'<t}} \exp\left(-\frac{(s^i(\mathbf{r}) - s^i(t'))^2}{2\sigma_{s^i}^2}\right) \quad (3)$$

where w is the height of the Gaussian potentials, τ_G is the rate of their deposition, and σ_{s^i} is the width of the Gaussian potentials associated with the collective variable s^i. The bias-exchange scheme involves swapping the atomic coordinates $\mathbf{r}^i$ and $\mathbf{r}^j$ of the two replicas i and j with an acceptance probability of

$$P_{ij} = \min\{1, \exp(\beta[V_G^i(s^i(\mathbf{r}), t) + V_G^j(s^j(\mathbf{r}), t) - V_G^i(s^j(\mathbf{r}), t) - V_G^j(s^i(\mathbf{r}), t)])\} \quad (4)$$

If the move is accepted, replica i continues with the configuration of replica j and vice versa. In other words, in BE-META each configuration evolves under a biasing potential that changes every time a move is accepted. This strategy can increase the diffusion capability of each replica in the collective-variable space, which, in turn, accelerates the exploration of the configuration space and improves the accuracy of the free-energy reconstruction.[51,66] If a neutral replica, a replica with no biasing potential added but allowed to participate in the conformational exchange, is included in BE-META and if the ratio w/τ_G is small, the neutral replica distribution can be considered an approximate canonical distribution.[51] In such a case, the probability distribution of the neutral replica follows the Boltzmann distribution and the free energy of a state i is defined as

$$F_i = -\frac{1}{k_B T}\log(P_i) \quad (5)$$

in which k_B is the Boltzmann constant, T is the absolute temperature, and P_i is the probability to observe the system in the state i.

2.2. Collective Variables (CVs). We employ seven CVs as listed below. WT-META only uses the first two CVs. BE-META has eight replicas: a neutral replica and seven biased replicas each using one of the seven CVs.

(1) *antibetarmsd* (PLUMED nomenclature[67]) counts the number of pairs of the 3-residue segments in the protein chain that are similar to an ideal antiparallel β block.[68] The ideal antiparallel β block is defined by the average structure of all antiparallel structures from the PDB database. The CV is evaluated as $s = \Sigma_\beta n[\mathrm{RMSD}(\{\mathbf{R}_i\}_{i\in\Omega_\beta}, \{\mathbf{R}^0\})]$, where $\mathrm{RMSD}(\{\mathbf{R}_i\}_{i\in\Omega_\beta}, \{\mathbf{R}^0\})$ is the root-mean-square deviation of the distance between pairs of atoms in the simulated protein with respect to the distance of the corresponding pairs of atoms in the reference structure, which is an ideal antiparallel in this case, Ω_β denotes the set of three continuous pairs of residues in the simulated protein, and $n \in [0, 1]$ indicates the similarity between each segment in the simulated protein and the ideal antiparallel β pattern as $n(\mathrm{RMSD}) = [1 - (\mathrm{RMSD}/0.1)^8]/[1 - (\mathrm{RMSD}/0.1)^{12}]$, where the RMSD (in nm) involves the distances among only five atoms including N, C_α, C, O, and C_β for each residue. In case of glycine, due to the missing of C_β, the corresponding H atom is used instead.[68]

(2) The *radius of gyration* is defined as $s = ([\Sigma_i^n|\mathbf{r}_i - \mathbf{r}_{COM}|^2]/[\Sigma_i^n m_i])^{1/2}$, where $\mathbf{r}_i$ and m_i are, respectively, the atomic coordinate and mass of the atom i, $\mathbf{r}_{COM}$ is the coordinate of the center of mass, and the sum is taken over all of the atoms in the peptide.

(3) The *number of* C_α *contacts* is defined by a continuous function

$$s = \sum_{i,j} \frac{1 - \left(\frac{d_{ij}}{r_0}\right)^6}{1 - \left(\frac{d_{ij}}{r_0}\right)^{12}} \quad (6)$$

where d_{ij} denotes the distance between the atoms i and j in the list of C_α atoms and r_0 is the cutoff distance of the contact that is chosen to be 0.65 nm for C_α contacts.

(4) The *number of* C_γ *contacts* is calculated by eq 6 with $r_0 = 0.5$ nm and the sum runs over the list of C_γ atoms.

(5) The *number of backbone hydrogen bonds*: Calculated using eq 6 with $r_0 = 0.25$ nm. The sums of i and j run over the lists of the backbone donors N and acceptors O, respectively.

(6) The *β-dihedral fraction* measures the similarity of the ψ dihedral angles to the reference value. The value of the CV is calculated as $s = (1/2)\Sigma_{i=1}^n(1 + \cos(\psi_i - \psi_i^{ref}))$, where the sum runs over all ψ dihedral angles of the peptide and $\psi_i^{ref} = 120°$ for a β structure.

(7) The *dihedral correlation* measures the similarity between adjacent ψ dihedral angles, $s = (1/2)\Sigma_{i=2}^n(1 + \cos(\psi_i - \psi_{i-1}))$ where the sum runs over all ψ dihedral angles of the peptide.

All replicas including the neutral replica were allowed to exchange every 100 ps. Gaussian potentials with height $w = 0.2$ kJ/mol were added to the time-dependent potential of the seven biased replicas every $\tau_G = 5$ ps. The width of the gaussians σ_{s^i} was 0.1, 0.05, 2.0, 1.5, 1.0, 0.5, and 0.5 for the CVs (1)−(7), respectively.

2.3. Computational Setup. Dodecahedral box of explicit water, in which the distance from any atom of the peptide to the box boundary is always greater than 1.8 nm, was used. Such a large distance was chosen to avoid interactions between the peptide and its periodic images when it is in an extended conformation. The dodecahedral box is chosen because its volume is only 0.71 times of a cubic box with the same periodic distance. The box contains 9726 TIP3P[69] water molecules. AMBER ff99SB force field[70] was employed together with recent corrections of the "ILDN" side-chain torsion parameters[38] and the helix−coil transition balance optimizations,[32] resulting the so-called ff99SB*-ILDN force field.[34] This combination of force field improvements has been shown to work well for a large variety of proteins including IDPs.[34,37,39] An excess ion concentration of 150 mM is set to reproduce physiological conditions, resulting in 32 Na^+ cations and 28 Cl^- anions and an overall charge neutrality. AMBER ff9X and their variants have been reported to facilitate ion crystallization due to the improper mixing of cation and anion parameters that causes the imbalance of cation−anion interactions.[71,72] Therefore, the ff99SB*-ILDN force field corrected by a recent ion reparametrization[73] was used for the Na^+ and Cl^- ions.

All simulations were performed with GROMACS 4.6[74] in the NpT ensemble. The PLUMED 1.3 plugin[67] was used for WT-META and BE-META. It is critically important to treat long-

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

range electrostatic interactions accurately;[75] hence, the Particle-Mesh Ewald method[76,77] was employed with a real space cutoff of 1.2 nm. The same cutoff was also used for the van der Waals interactions. The time-step was 2 fs. Temperature was kept constant at 310 K using the Bussi−Donadio−Parrinello velocity rescale algorithm[78] and pressure at 1 atm using the Parrinello−Rahman barostat.[79]

2.4. Analysis Methods. Cluster analysis, dihedral angle calculation, root-mean-square deviation measurement, number of backbone N−O contacts, and secondary structure content analysis were performed. Additional analyses include the calculation of chemical shift index, dynamic secondary structure content assignment and secondary-structure-based free energy calculation.

2.4.1. Chemical Shift Index. Chemical shift index (CSI) is a technique used to identify protein secondary structure based solely on backbone NMR chemical shift data. Originally CSI technique was developed for the analysis of $^1H_\alpha$ chemical shifts.[80] The technique was then extended to include the $^{13}C_\alpha$, $^{13}C_\beta$, and carbonyl ^{13}C chemical shifts.[81] CSI is a ternary index adopting the values of −1, 0, and 1. CSI of each of the four atoms is assigned on the basis of the chemical shift values of the atoms. For each amino acid a specific range of chemical shift of each atom is determined. CSI of an atom is equal to 1 if the measured chemical shift of that atom is greater than the specified range. CSI takes the value of −1 if the measured chemical shift is less than the range. CSI is equal to 0 if the measured chemical shift is within the range. The secondary structure is then established by the following rules:[81] an α-helix is defined when four or more "−1" C_β/H_α CSIs and/or "1" C_α/C CSIs are sequentially found, a β-strand is defined when three or more "+1" C_β/H_α CSIs and/or "−1" C_α/C CSIs are sequentially found, all other regions are considered coil. When chemical shifts of three or more atoms are available for the same residue, a consensus CSI algorithm based on a simple "majority rules" approach (i.e., two out of three, or three out of four) allows an improved prediction of secondary structure.[81] The overall performance of the consensus predictions can be as high as 92%.[81]

In our analysis, the chemical shifts of the C_α, C_β, carbonyl C, and H_α atoms were predicted with the ShiftX software.[82] The consensus CSI algorithm combining the CSI values of these four atoms for each residue were then evaluated using the CSI software.[81] The secondary structure was finally determined as an α-helix if four or more positive consensus CSI values are sequentially found, a β-strand if three or more negative consensus CSI values are sequentially found, a coil for all other cases.

2.4.2. Dynamic Secondary Structure Content. The secondary structure content is analyzed using DSSP (Dictionary of Protein Secondary Structure).[83,84] Based on the hydrogen bond patterns and geometric properties, each residue is assigned one of eight letter codes, G for 3_{10} helix, H for α helix, I for π helix, T for hydrogen bonded turn, E for extended β-strand, B for β-bridge, S for bend, and C or ∼ for random coil. The discrete assignment does not necessarily describe the continuum features of thermal fluctuations that are expected to play an important role in IDP structural flexibility. A method called DSSPcont has been proposed to take into consideration the dynamics of a static structure by calculating a weighted average over 10 discrete DSSP assignments with different hydrogen bond thresholds for a single structure.[85,86]

In our analysis, to take into account the protein dynamics driven by thermal fluctuations, we averaged over 10 discrete DSSP assignments of 10 adjacent frames with a time-step of 5 ps. The resulted average structure has each residue assigned with a percentage likelihood of all eight DSSP secondary structure states, thus representing a dynamic structural content.

2.4.3. Secondary-Structure-Based Free-Energy Calculation. Since the free energy landscape is a multidimensional surface, reconstructing the free energy as a function of one or few CVs does not usually provide a satisfactory description of all the relevant states and transitions. Moreover, free-energy barriers and minimum free-energy paths can change drastically in different representations. As will be shown in the next sections, all seven CVs in our BE-META have important contributions to the exploration of the conformational space. Since we are interested in conformations, we propose a free-energy calculation purely based on the probability distribution of the secondary structure content. This postprocessing protocol is applicable to the canonical ensemble including the neutral replica of BE-META.

Secondary-structure-based free-energy method involves calculating the free energy of a conformational ensemble represented by a certain representative structure, which is referred to as a reference structure hereafter. First of all, the dynamic secondary structure content of each frame in a canonical trajectory is calculated as described in Section 2.4.2. Next, in terms of secondary structure contents we define a reference structure representing a conformation or a group of conformations whose free energy is to be calculated. Then, each frame of the trajectory is weighted by a number representing the similarity of its dynamic secondary structure and the chosen reference structure given by

$$w(t) = \prod_{i=1}^{n} \sum_{s=1}^{8} p_{i,s}(t) c_{i,s}^{\mathrm{ref}} \tag{7}$$

where the product runs over all residues, the sum runs over all 8 DSSP secondary structure types, $p_{i,s}(t)$ denotes the probability for the residue i to adopt the structure s at time t, and $c_{i,s}^{\mathrm{ref}}$ is a switching factor performing a discrete counting with a value of 1 when residue i in the reference structure adopts structure s and 0 otherwise. For instance, in random-coil conformations, each residue can adopt one of the three secondary structures including coil (∼), turn (T), and bend (S) with the same probability; therefore, the switching factors counting these three states in the reference structure representing the random-coil ensemble, that are, $c_{i,\sim}^{\mathrm{coil}}$, $c_{i,\mathrm{T}}^{\mathrm{coil}}$, and $c_{i,\mathrm{S}}^{\mathrm{coil}}$, are assigned with the value of 1 while the other switching factors of other secondary structures are assigned with the value of 0. The total occurrence of the reference structure in the trajectory is then defined as the sum of all weights along the trajectory

$$P_{\mathrm{ref}} = \sum_{t} w(t) \tag{8}$$

For the canonical ensemble, the free energy of the reference state is evaluated by eq 5. The free-energy difference between two conformational ensembles described by two reference states ref 1 and ref 2 is thus given by

$$\Delta F = F_{\mathrm{ref2}} - F_{\mathrm{ref1}} = -\beta \log\left(\frac{P_{\mathrm{ref1}}}{P_{\mathrm{ref2}}}\right) \tag{9}$$

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

3. RESULTS AND DISCUSSIONS

3.1. Sampling Efficiencies by Different Simulation Methods. To find a simulation protocol that is suitable to explore IDP conformational space, we employ unbiased MD, WT-META, and BE-META starting from the exact same random-coil structure of the above-mentioned Neh2 peptide. Both the unbiased MD and WT-META reached 3 μs, and each of the eight replicas of BE-META covered 460 ns making up a total of 3.68 μs for BE-META. To have a fair comparison, only the first 375 ns of each BE-META replica (total of 3 μs) is considered in this section. The computational effort to reach 3 μs is comparable with each of the methods.

Figure 1 shows the time dependence of the number of clusters explored by each method. Different Neh2 structures explored along each simulation are grouped into clusters[87] with an RMSD cutoff of 0.2 nm. The number of clusters reaches a stable value after approximately 1 μs in all cases. Beyond 1 μs, the averaged numbers of explored clusters by unbiased MD, WT-META, and BE-META are 331 ± 24, 520 ± 27, and 548 ± 21, respectively. With a comparable computational effort, both the WT-META and BE-META have a considerably higher sampling efficiency than the unbiased MD. It is also worth noticing that after 1 μs, BE-META can explore slightly more clusters than WT-META although they are within the margin of error.

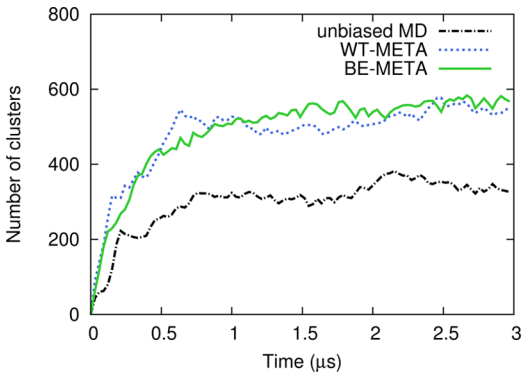


Figure 1. Total number of clusters explored as a function of time. After approximately 1 μs, the number of clusters roughly reaches a stable value. WT-META (dotted blue line) and BE-META (solid green line) have higher sampling efficiency compared to unbiased MD (dashed-dotted black line).

The same feature can be observed in the Ramachandran plot for the ϕ and ψ backbone dihedral angles of all Neh2 residues obtained from the three simulations (see Figure 2). Qualitatively, the possible combinations of the dihedral pairs including rare structures are extensively explored in all cases. However, the visited regions in both the WT-META and BE-META appear to be larger and more connected than those in the unbiased MD.

More convergence indicators such as the time dependence of hill heights in WT-META as well as the time evolution of the reconstructed free-energy profiles as functions of the biased CVs in both WT-META and BE-META are shown in the Supporting Information (see Figures S1−S10).

3.2. Reproduction of the Bound-State-Like Conformation of Neh2. *3.2.1. Diversity of the Conformational Sampling.* To assess the effectiveness of sampling, we calculate the distribution of the C_α RMSD between the Neh2 structures (residues PHE71 to LEU84) sampled along the simulation with respect to the available X-ray structure of the Neh2 peptide in complex with the Kelch domain of Keap1 (pdb code: 2flu[59]). The results are shown in Figure 3. The data set of WT-META

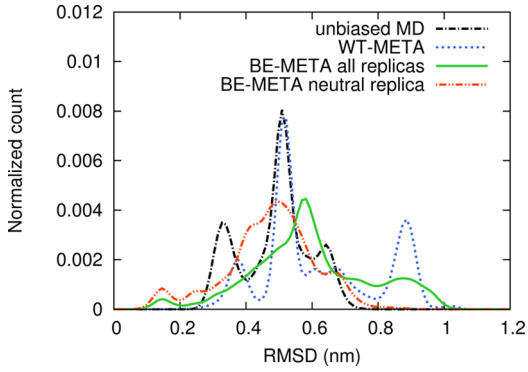


Figure 3. Normalized histogram of the RMSD between the modeled Neh2 conformations with respect to the X-ray structure (pdb code: 2flu[59]). The histograms obtained from unbiased MD, WT-META, BE-META, and neutral replica of BE-META are shown with the dashed-dotted black line, dotted blue line, solid green line, and dashed-dotted red line, respectively. The unbiased MD appear as the most sampling-insufficient while the BE-META simulation and especially its neutral replica emerge as the most sampling-effective.

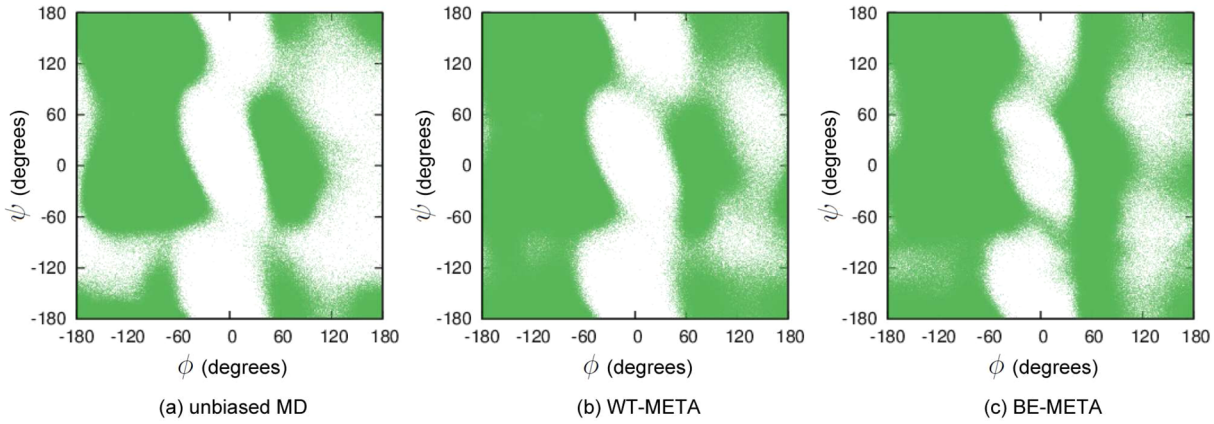


Figure 2. Ramachandran plots of the ϕ and ψ backbone dihedral angles of all Neh2 residues obtained from (a) the unbiased MD, (b) WT-META, and (c) BE-META. The data are plotted every 5 ps along each simulation. Qualitatively, both the WT-META and BE-META can explore better the (ϕ, ψ) space than the unbiased MD.

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

and BE-META with all replicas are properly reweighted to remove the bias effect using the reweighting technique introduced in refs 88 and 89. For the unbiased MD (dashed-dotted black line), the RMSD values fall between 0.24 and 0.83 nm and the plot has a sharp peak close to 0.5 nm. The sharp peak can be interpreted as a trap in the local free energy minima and insufficient sampling, that can be expected from an unbiased MD simulation. The dotted blue line shows the RMSD distribution in case of the WT-META simulation, that is, between 0.18 and 1.13 nm. The WT-META simulation samples a larger range of RMSD values than the unbiased MD. However, WT-META also experiences a sharp peak close to 0.5 nm as in unbiased MD. Interestingly, WT-META explores better the conformational space with large RMSD values (0.8−1 nm) with respect to the X-ray structure.[59] The RMSD distribution obtained from BE-META (solid green line) spans over the largest interval of RMSD values among the three cases. The BE-META simulation can access structures with RMSD as low as 0.05 nm with respect to the X-ray reference structure. Some of the structures explored by BE-META also occupy the larger RMSD region beyond 1.13 nm. An interesting feature emerges when the RMSD distribution of the neutral replica of BE-META (dash-dotted red line) is analyzed: The neutral replica can explore the low RMSD region close to the lower bound of the three cases. Moreover, the neutral replica of 375 ns can sample a wider range of RMSD values compared to the 8-time-longer unbiased MD of 3 μs. Overall, BE-META and its neutral replica prove to be the most sampling-effective case.

3.2.2. Reproduction of the Key Backbone Contacts in the Bound-State-Like Structure of Neh2. Backbone Contacts Observed in the X-ray Structure. A β-hairpin conformation is defined by the hydrogen bonds between the NH and CO groups in the backbone of a protein.[90] We found a total of 8 contacts between all backbone N and O atoms in the bound-state-like X-ray structure of the Neh2 peptide[59] (Table 1 and Figure 4). The

Table 1. Backbone N−O Contacts of the X-ray Neh2 Structure[59]

residue@atom	residue@atom	distance (nm)
GLN75@N	LEU84@O	0.28
GLN75@O	LEU84@N	0.30
ASP77@N	GLU82@O	0.28
ASP77@O	GLU82@N	0.42
ASP77@O	GLU79@N	0.32
ASP77@O	THR80@N	0.36
ASP77@O	GLY81@N	0.29
THR80@O	GLU82@N	0.33

GLN75@O-LEU84@N contact (0.30 nm) and ASP77@N-GLU82@O contact (0.28 nm) facilitate the β-bridge formation of the residues LEU76 and PHE83. On one side of the β-bridge, the residues ASP77 and GLU82 do not form an extended β-strand due to the rather weak ASP77@O-GLU82@N contact (0.42 nm). On the other side of the β-bridge, the residues GLN75 and LEU84 are observed with another strong backbone contact, that is, the GLN75@N-LEU84@O contact of 0.28 nm. However, the X-ray structure of Neh2 was truncated at residue LEU84, and therefore, an extended β-strand of GLN75 and LEU84 was not reported.[59] The hairpin loop of Neh2 has an interesting tight turn structure, in which the O atom of the residue ASP77 is observed to have strong contacts with the N atoms from other loop residues including GLU79 (0.32 nm),

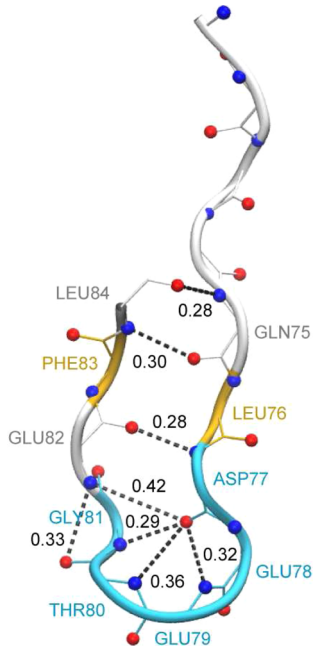


Figure 4. Backbone representation of the Neh2 X-ray structure.[59] The coil, β-bridge, and turn structures are shown in light gray, yellow, and cyan, respectively. The residue names are shown beside the residues in the same color codes. The backbone N and O atoms are represented as spheres in blue and red, respectively. The backbone N−O contacts are shown with dashed lines and the numbers associated with the lines indicate the contact distances (in nm).

THR80 (0.36 nm), and GLY81 (0.29 nm). In the X-ray structure, the residues GLU78, GLU79, and THR80 are found in a turn conformation while the residues ASP77, GLY81, and GLU82 remain as parts of a random coil.

Backbone Contacts Observed in Simulations. The backbone N−O contacts found in simulations are shown in Figure 5 in a form of a contact map between each and every Neh2 residue. The number of backbone N−O contacts between each pair of residues is averaged over time starting from the convergence of 1 μs in each simulation. The data set of WT-META and BE-META with all replicas (Figure 5b and c) are properly reweighted to remove the bias effect using the reweighting technique introduced in refs 88 and 89. Since the N_i-O_i and N_i-O_{i-1} contacts, that is, within the same and adjacent residues, respectively, are not involved in secondary-structure recognition, they are not shown in the map to avoid confusion. The yellow square in each map defines the loop region (i.e., from residue ASP77 to GLU82) of the β-hairpin Neh2 as seen in the X-ray structure.[59] Next, the three methods are compared in detail.

In the unbiased MD simulation (Figure 5a), the bound-state contacts between the loop residues including ASP77-GLU79, ASP77-THR80, and THR80-GLU82 are sampled. However, none of the contacts facilitating the experimental β-bridge are observed. Similarly, in the WT-META simulation (Figure 5b), we also find three contacts between the loop residues, including ASP77-GLU79, ASP77-THR80, and THR80-GLU82, that resemble the contacts in the bound-state structure.[59] The two β-facilitating contacts, that is, GLN75-LEU84 and ASP77-GLU82, are not found. Therefore, there is no β-hairpin similar to the X-ray structure during the WT-META simulation. In the BE-META simulation and its neutral replica (Figure 5c and d), we find all loop N−O contacts reported in the X-ray structure[59] including the ASP77-GLU79, ASP77-THR80, ASP77-GLY81,

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

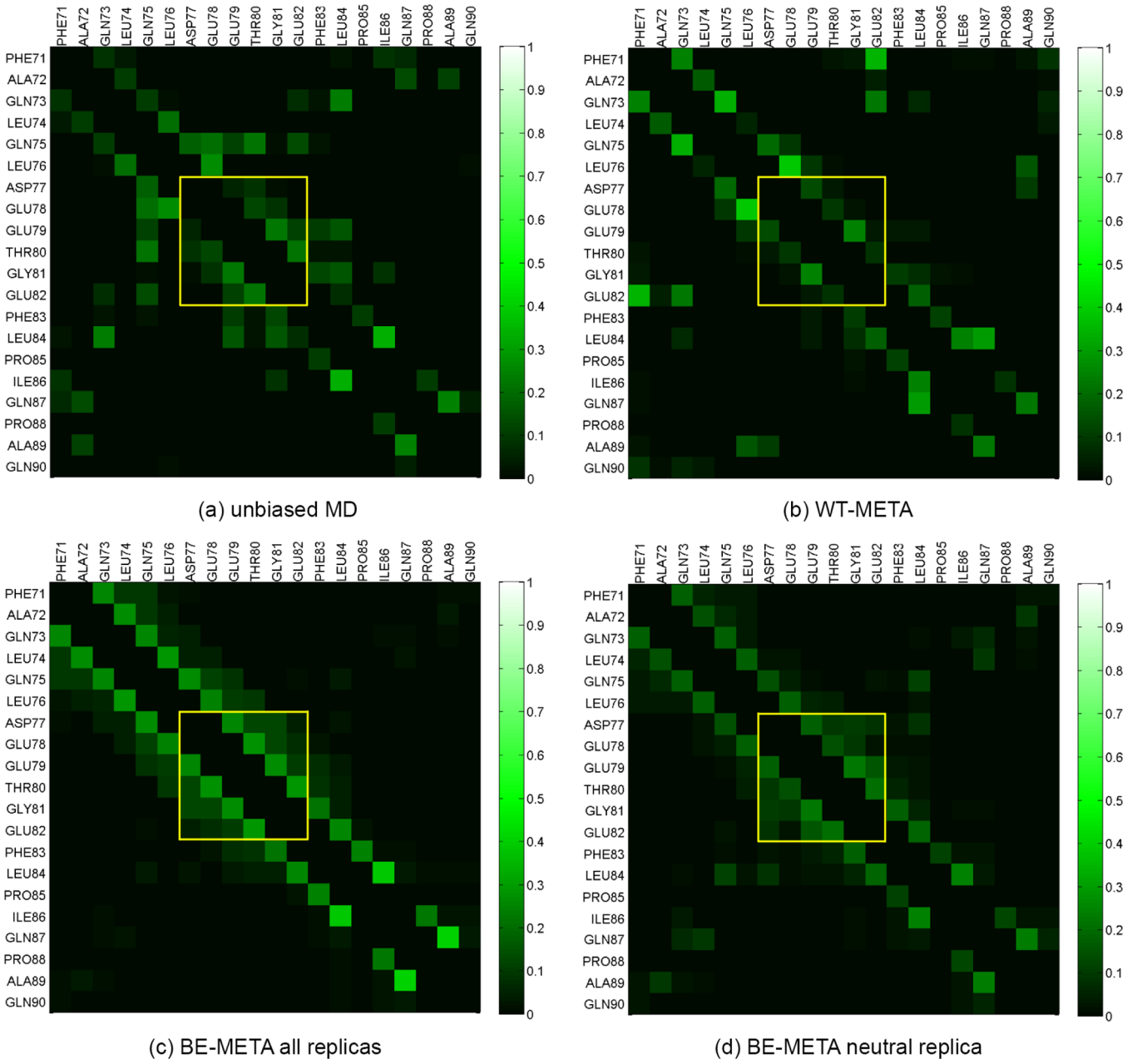


Figure 5. Hydrogen-bond maps obtained by time averaging starting from 1 μs in (a) unbiased MD, (b) WT-META, (c) BE-META, and (d) the neutral replica of BE-META. The yellow square represents the loop region (from ASP77 to GLU82) of the β-hairpin Neh2 as seen in the X-ray structure.[59] Only BE-META and its neutral replica reproduce well the X-ray native contacts, including the ASP77-GLU79, ASP77-THR80, ASP77- GLY81, and THR80-GLU82 contacts.

and THR80-GLU82 contacts. In the neutral replica, we observe the crucial β-facilitating contacts including GLN75-LEU84 and ASP77-GLU82. These are clear signs for the reproduction of the bound-state Neh2 conformation by the neutral replica of the BE-META simulation. Interestingly, as we predicted, the β-hairpin can also be extended. In addition, the neutral replica can explore several other β-hairpin conformations in which the hairpin loop is shifted with 1, 2, and 3 residues to both sides of the hairpin loop in the X-ray structure. Moreover, several 3_{10}, α, and π helical conformations with the contact patterns of $i-(i+3)$, $i-(i+4)$, and $i-(i+5)$ are also discovered in the neutral replica. These helical conformations support our hypothesis that the hairpin loop can also adopt a helical structure at times.

Our results show that while both the WT-META and BE-META have comparable sampling efficiencies (i.e., number of explored conformations), only BE-META with a larger set of CVs stands out for its sampling effectiveness. BE-META proves to be an effective enhanced-sampling protocol that fits well for studies of IDP ensembles. The chosen set of CVs for BE-META proves to be qualitatively sufficient for the exploration of not only β-strand but also helical propensities in spite of the fact that there is no biasing CV explicitly related to helical structures.

The probability of successful exchange between the neutral replica and the other biased replicas of BE-META is lower than 12% during the course of a 460 ns simulation for each replica. A small percentage is often the result of a small w/τ_G ratio.[51] In such a case, the neutral replica can be considered a good

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

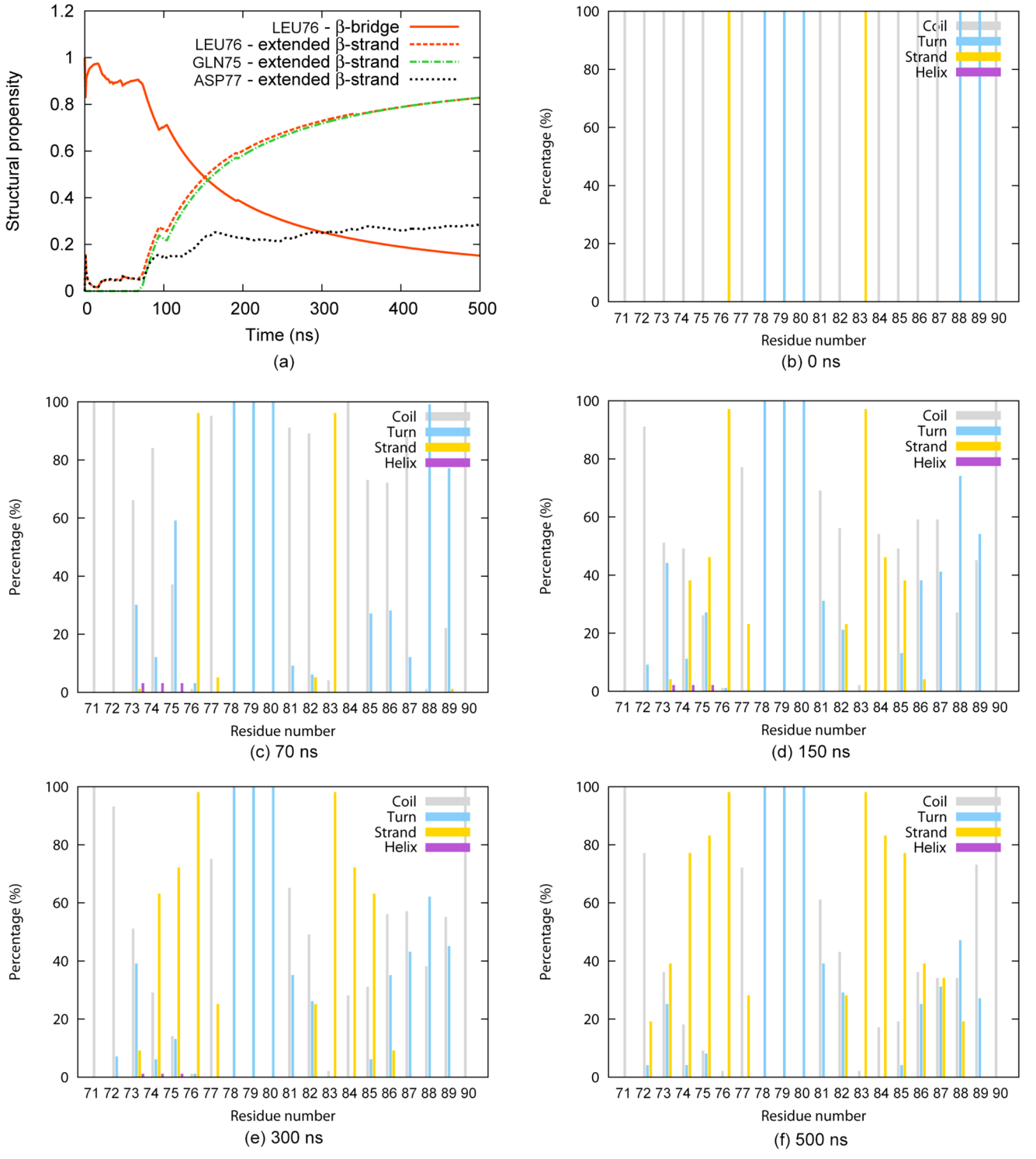


Figure 6. (a) Cumulative moving average of the β propensity in the residues GLN75 (dashed-dotted green line), LEU76 (solid and dashed red lines for β-bridge and extended β-strand respectively), and ASP77 (dotted black line). (b, c, d, e, and f) show the structural propensities for each residue at 0, 70, 150, 300, and 500 ns, respectively. These are the important points where critical structural changes happen. The bound-state-like structure is not stable beyond 70 ns.

approximation of the canonical ensemble. Its probability distribution thus follows the Boltzmann distribution and the free energy of a state is given by eq 5. Due to the comparable sampling efficiency and effectiveness between the neutral replica and the whole BE-META simulation, hereafter, we can confidently analyze the neutral replica as a sampling representative of BE-META.

3.2.3. Occurrence, Stability, and Dynamical Properties of Bound-State-Like Conformation. Structural analysis is performed using the DSSP pattern recognition of hydrogen bonds and geometric features.[37,83,84] This method assigns the X-ray structure of the 14-residue Neh2 peptide with a sequence of structural codes, that is, ~ ~ ~ ~ ~ B ~ TTT ~~ B ~, in which ~, B, and T encode for coil, β-bridge, and turn, respectively. We

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

also perform a DSSP analysis on the structure at every 5 ps of the 460 ns neutral replica assigning DSSP structural codes to a total of 92000 structures. We find 76 times in which the *exact* X-ray sequence occurs among all 92 000 sequences. The bound-state-like structure is explored for the first time at around 338 ns of the neutral replica. The 76 structures sharing the same structural contents with the X-ray structure have the C_α RMSD at the region from LEU76 to PHE83 with respect to the X-ray structure ranging from 0.02 to 0.07 nm. The neutral replica of BE-META can reproduce the experimental structure *without any* direct guidance toward this structure during the simulation. It is worth mentioning that the exact X-ray DSSP sequence occurred in neither unbiased MD nor WT-META during 3 μs simulations. These findings again prove the better sampling quality of BE-META over unbiased MD and WT-META. More importantly, they suggest that the folded bound-state-like structure is a rare folded conformation of Neh2, whose stability needs to be further investigated.

A 500 ns unbiased MD simulation was then performed on the structure with the lowest RMSD with respect to the X-ray structure among the 76 above-mentioned exact structures. Figure 6a shows the cumulative moving average of the β propensities of different residues in the MD simulation. The β-bridge propensity of residue LEU76 with a solid red line and the extended β-strand propensity of the same residue with a dashed red line. The extended β-strand propensities of residues GLN75 and ASP77 are also shown in dashed-dotted green and dotted black lines, respectively. The β partners of these residues, that is, PHE83, LEU84, and GLU82, respectively, have the same β propensity as their partners, and thus are not presented in this plot. During the first 70 ns of the unbiased MD simulation, GLN75 does not have any β propensity. LEU76 still highly favors a β-bridge content as in its starting configuration. However, LEU76 does not always stay in a β-bridge conformation. At times LEU76 and one of its adjacent residues, ASP77, can adopt an extended β-strand structure. After the first 70 ns of the simulation, the peptide structure changes drastically. The β-bridge propensity of LEU76 decreases and its extended β-strand propensity increases simultaneously at an exponential rate. Concurrently, the extended β-strand propensity of GLN75 increases almost at the same pace as that of LEU76. Interestingly, the extended β-strand propensity of ASP77 slowly increases and reaches a stable state at around 300 ns. Figure 6(b, c, d, e, and f) shows the detailed structural propensity for each residue at 0, 70, 150, 300, and 500 ns, respectively. Our 500 ns unbiased MD simulation reveals that the bound-state-like conformation of the free Neh2 peptide is not stable beyond 70 ns. The free Neh2 peptide prefers adopting a more stable extended β-strand involving at least two more pairs of residues, i.e., GLN75-LEU84 and ASP77-GLU82. The probabilities of observing the pairs GLN75-LEU84 and LEU76-PHE83 are equivalent after 350 ns. They are both three times higher than the probability of observing the pair ASP77-GLU82. Thermal fluctuations still allow the peptide to revisit its starting structure with the β-bridge at LEU76-PHE83 with a probability of 15% after 500 ns. Our simulation also indicates that the loop containing the residues GLU78, GLU79, and THR80 form an absolutely stable turn structure during 500 ns.

3.2.4. Free Energy of the Bound-State-Like Conformation of Neh2. We employ the method described in Section 2.4.3 to calculate the free-energy difference between the bound-state-like and the random-coil ensembles. In terms of secondary-structure representation the bound-state-like structure is described by the sequence ~ ~ ~ ~ ~ B ~ TTT ~~ B ~ and the random-coil ensemble contains all the conformations in which each residue can adopt one of the three structures including ~, T, and S. The free-energy calculation method is based on the probability distribution of a state with a defined secondary structure, and hence is CV-independent. Figure 7 shows the time evolution of

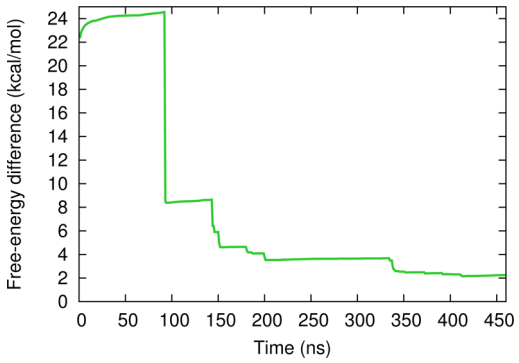


Figure 7. Time dependence of the free-energy difference between the bound-state-like and the disordered conformations of Neh2 estimated from the neutral replica. The free-energy difference converges to 2.3 ± 0.1 kcal/mol, that is, almost four times of the thermal energy; the bound-state-like structure is not reachable from a disordered structure if the system undergoes thermal fluctuations only.

the free-energy difference between the bound-state-like and random-coil conformations, which converges to 2.3 ± 0.1 kcal/mol. The free-energy difference and the error are the mean and standard error, respectively, calculated for the last 110 ns of the neutral replica starting from 350 ns when the plot shows a good convergence. The free-energy difference between the bound-state-like and the disordered conformations is almost four times of the thermal energy. That explains why this folded structure is not reachable with unbiased MD.

3.3. Reproduction of the Extended β-Hairpin Conformation of Neh2 in Its Free State. *3.3.1. Structural Propensities.* Figure 8 shows the structural propensity of each Neh2 residue as explored by the neutral replica. The first two residues at each terminal prefer the random coil state. The rest of

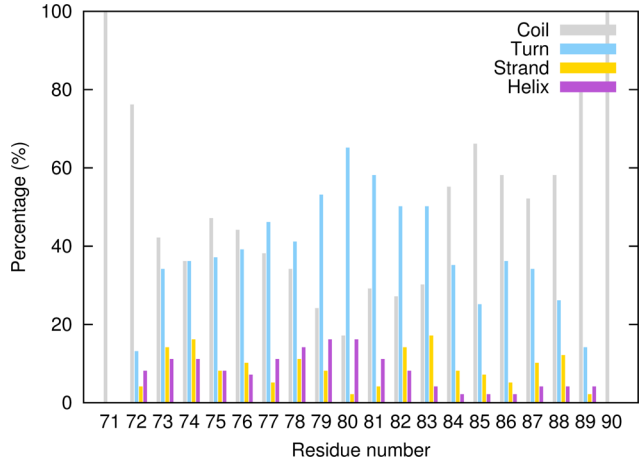


Figure 8. Secondary structure contents for each residue sampled by the neutral replica. The coil, turn, strand, and helix contents are shown in gray, cyan, yellow, and purple, respectively. The turn propensity dominates from ASP77 to PHE82 while the loop propensity dominates for the rest of the residues. This shows the intrinsically disordered nature of the Neh2 peptide.

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

the residues have large probabilities (>36%) to be in either turn or random coil conformations. There is a higher turn propensity compared to coil at the middle residues, that is, from ASP77 to PHE83. This region was identified as the hairpin loop region with a tight turn conformation in the X-ray structure. Interestingly, the neutral replica can sample both β-strand and helical content at each residue except for the terminal ones. However, the chance for each Neh2 residue to be in these two well-defined secondary structure is small (<20%). This agrees with the intrinsically disordered nature of Neh2. The five residues from ASP77 to GLY81 have a higher helical propensity than β-strand propensity. This finding verifies our hypothesis of a helical portion in this region, as previously seen in the N−O contact map (Figure 5d). The rest of the residues have a higher β-strand propensities compared to helical. The highest β trends are at PHE83 (17%), LEU74 (16%), GLU82 (14%), and GLN83 (14%). It is noteworthy that residue PHE83 was found to adopt a β-bridge structure together with residue LEU76 in the X-ray structure. Our simulation also discovers a β trend at LEU76, but its occurrence (10%) is not as high as that of PHE83 (17%). Most of the missing residues from X-ray experiment (i.e., from PRO85 to PRO88) are found with a noticeable β trend (5−12%). This confirms our hypothesis that the β-sheet conformation of Neh2 can be extended beyond the experimentally explored β-bridge if a longer Neh2 chain is considered.

Our findings prove that the bound-state-like structure of Neh2 is only a transient conformation in its free state. Other conformations with more extended β-strands appear to be explored with higher probabilities. In the following sections, we further compare the structural properties of these structures with available NMR data of Neh2 in its free state.[58]

3.3.2. Qualitative Agreement in the Secondary Structure Contents Predicted by the Chemical Shift Indices. Figure 9 shows the average consensus chemical shift indices (CSIs) for each Neh2 residue in the neutral replica simulation (see Section 2.4.1 for details of the calculation). Consecutive negative CSIs define a β-strand content at two separate regions, ALA72-GLU78 and GLU82-GLN87. Four consecutive slightly positive CSIs define an slight α-helix propensity at GLU78-GLY81. This finding is in qualitative agreement with the NMR observation, which was also based on CSI evaluation.[58] Both the NMR experiment and simulation suggest that the Neh2 peptide in its free state is intrinsically disordered with local elements of secondary structure, that is, a hairpin structure with an antiparallel hydrophobic β-sheet involving up to 6 pairs of residues and a hydrophilic loop adopting α-helix structures at times.

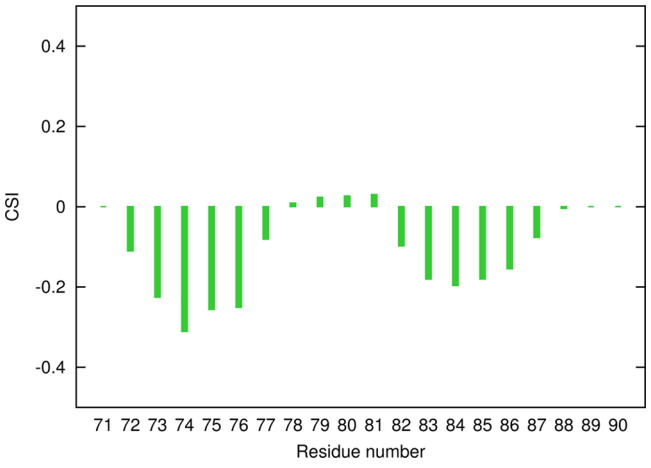


Figure 9. Consensus chemical shift index (CSI) for each residue of the free Neh2 peptide in the neutral replica simulation. Consensus CSI values are calculated from the combined CSIs of the C_{α}, C_{β}, carbonyl C, and H_{α} atoms. Four or more consecutive positive CSIs indicate an α-helix, three or more consecutive negative CSIs indicate a β-strand, all other cases indicate a coil structure.

3.3.3. Reproduction of ^{1}H, ^{1}H Nuclear Overhauser Effects. A ^{1}H, ^{1}H nuclear Overhauser effect (^{1}H, ^{1}H NOE) is observed when two protons are within 0.6 nm in distance. The NMR experiments found weak-to-moderate, but evident, ^{1}H, ^{1}H NOEs between residues within the two β-strands. This finding supports the hypothesis that the two β-strands fold into a small antiparallel β-sheet with a hydrophilic hairpin loop at the conserved EETGE motif.[58] In our simulation, we also find evidence of such a β-hairpin conformation. The reproduction of the proton pairs within 0.6 nm is presented in Table 2. Our simulation reproduces 15 of the 23 pairs of proton showing ^{1}H, ^{1}H NOEs in NMR experiment.

Table 2. Pairs of Protons Showing ^{1}H, ^{1}H NOEs (i.e., within a Distance Smaller than 0.6 nm)[a]

region	residue	proton	residue	proton
β-sheet	LEU74	Hδ	PRO85	Hα
			LEU84	(HN)
		Hϵ	PRO85	(Hα)
	GLN75	HN		(Hα)
			LEU84	HN
		Hβ		HN
		Hγ		HN
	LEU76	HN	PHE83	Hδ, (Hϵ)
			LEU84	(HN)
		Hδ	PHE83	(HN), Hδ, Hϵ
			GLU82	HN
			GLY81	(HN)
	ASP77	HN	LEU84	(Hδ)
			GLU82	HN
		Hβ		HN
loop	GLU78	HN	GLU79	HN
		Hα	THR80	HN
	GLU79	HN		HN
	THR80	HN	GLY81	HN
	GLY81	HN	GLU82	HN

[a]Data are adopted from Table 1 in ref 58. The pairs with the second proton in the parentheses are not observed in our simulation. The rest are reproduced by our simulation.

Our simulation also observes more ^{1}H, ^{1}H NOEs between residues in the β-sheet and loop regions as indicated by NMR experiment. Figure 10 is a distance map in which the distances of all possible pairs of hydrogen atoms are averaged out for the residues and shown in a color scale from dark green (≤0.2 nm) to white (≥0.7 mm). Weak-to-moderate evidence of different β-sheet patterns is found.

3.4. Other Highly Probable Conformations of Neh2. A cluster analysis based on the RMSD of the peptide mainchain with an RMSD cutoff of 0.2 nm was performed on the neutral replica of BE-META to identify the ten most probable conformations of the Neh2 peptide. An averaged secondary structure content was calculated for each conformation. The free energy difference of each conformation with respect to the

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

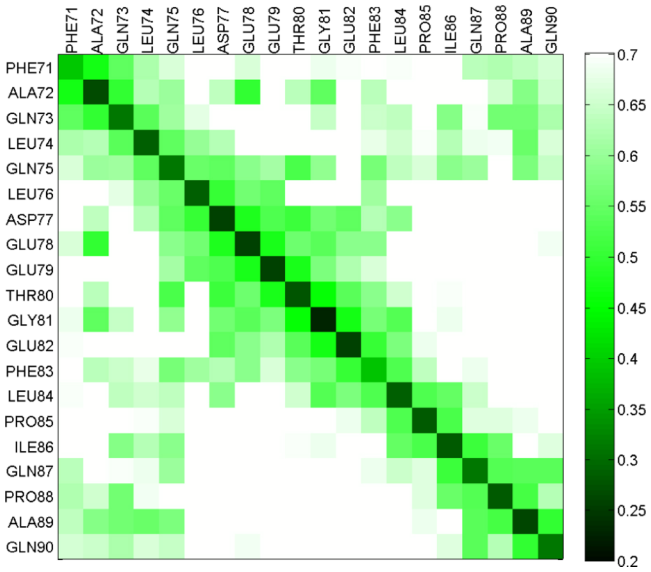


Figure 10. Time-average cross-residue distance in the neutral replica simulation. Distances are shown with a color scale from dark green (≤0.2 nm) to white (≥0.7 nm).

random coil state was then estimated (Section 2.4.3). Figure 11 shows the averaged free energy differences of the ten most

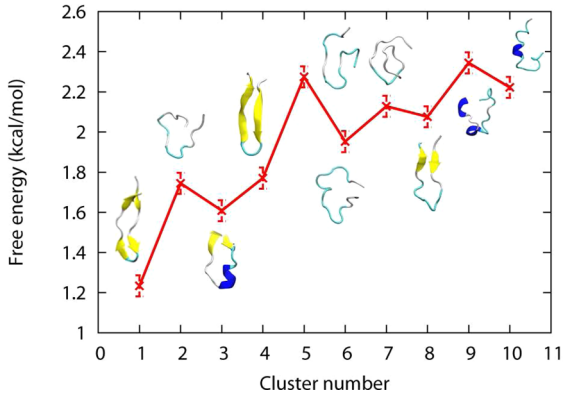


Figure 11. Free energy difference with respect to the random coil state of the ten most probable clusters explored by the neural replica. Light gray, cyan, yellow, and blue represent the coil, turn, β-strand, and α-helix structures, respectively. The fifth cluster, that has higher free energy than most clusters, resembles the bound-state-like conformation of the Neh2 peptide.

adopted conformations explored by the neutral replica of BE-META. The average and the standard error were calculated over the last 110 ns of the neutral replica. The fifth conformation appears to match the bound state of Neh2 with a similar free energy value (2.3 ± 0.1) kcal/mol. Most conformations have lower free energy than the bound-state-like conformation. The forth conformation resembles the extended β-hairpin structure of Neh2 with a free energy of 1.8 ± 0.1 kcal/mol compared to the random coil state.

4. CONCLUSIONS

We have computationally examined a 20-mer peptide derived from the N-terminal Neh2 domain of the Nrf2, a key transcription factor regulating the oxidative stress response.[57] Neh2 has been characterized as intrinsically disordered with a weak but non-negligible tendency of a β-hairpin formation in its free state.[58] Experimental data also shows that when binding to the Kelch domain of Keap1, Neh2 forms a short but stable β-hairpin.[59] The disordered nature of Neh2 challenges both the experimental and computational characterizations of its conformational space. In previous studies, a few long unbiased MD and short enhanced-sampling simulations have been performed for a 16-mer derivation of Neh2.[37,91,92] In these studies, short and extended β-hairpin conformations were explored with missing information on their corresponding free energies and stabilities. Here, we have explored the conformational space of the extended 20-mer Neh2 peptide using unbiased MD and two metadynamics variants, namely well-tempered metadynamics[60] and bias-exchange metadynamics,[51] with increased sets of relevant CVs. All simulations reached 3 μs and beyond, yielding the first extensive enhanced conformational sampling of Neh2.

We first showed that metadynamics provides a significantly better conformational sampling compared to unbiased MD of the same computational expense and time. This is reflected in the number of observed clusters, the exploration of backbone dihedral space, and the RMSD distribution with respect to the reference bound-state-like structure.[59] Among the two metadynamics variants employed in this study, BE-META with a larger set of CVs proves to be a more effective sampling method suitable for exploring the complex conformational space of Neh2. Indeed, the adding of five more relevant CVs (total of seven) helped to observe the rare bound-state-like conformation that were not reproduced in either unbiased MD or WT-META, and has never been found in any other computational study of free Neh2. Additional unbiased MD simulations suggested that the bound-state-like conformation is unstable when Neh2 is in its free state. The conformational ensemble explored by BE-META shows that the bound-state-like structure is a rare structure with a free-energy difference of 2.3 ± 0.1 kcal/mol with respect to the random coil ensemble. This is almost four times the thermal energy, making it challenging to fully sample the transition from the random coil structure to the unfavorable bound-state-like conformation of Neh2 and vice versa.

Besides reproducing the bound-state-like structure,[59] BE-META also explores the extended β-hairpin conformation of Neh2 that has been observed in its free state.[58] Our predictions of chemical shift indices and signals for the occurrence of ^{1}H, ^{1}H NOEs compared well with available NMR data.[58] The free energy of the observed extended β-hairpin conformation is 1.8 ± 0.1 kcal/mol higher than that of the random coil ensemble.

In addition to the free- and bound-state-like structures, our simulation also obtains other structural propensities. The chosen set of CVs is general and uniformly applied to all Neh2 residues. They do not require advanced knowledge of experimentally observed folded structures, and thus do not bias toward a limited set of predefined structures: all physically relevant conformations are explored. Moreover, these CVs appear to be qualitatively sufficient for the reproduction of not only β-sheet but also helical contents despite the indirect relation of these CVs to helical structures. These features make this protocol highly applicable for exploring the IDPs' conformational space even when experimental guidance or folded structures are unavailable.

To the best of our knowledge, this is the first extensive computational study and comprehensive description of the vast conformational ensemble of Neh2, yielding not only the experimentally observed β-hairpin structure but also the rare

dx.doi.org/10.1021/ct5004803 | J. Chem. Theory Comput. 2014, 10, 5081−5094

bound-state-like and other conformations, that have not been observed elsewhere.

■ ASSOCIATED CONTENT

Supporting Information

Convergence indicators of WT-META and BE-META simulations, including the time dependence of hill heights in WT-META and the evolution of the free-energy profiles in both WT-META and BE-META. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mkarttu@gmail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Elio Cino for useful discussions. We also thank Compute Canada for computational resources. Financial support was provided by the Canadian Institute of Health Research (CIHR; MOP 74679), University of Waterloo, and Natural Sciences and Engineering Research Council of Canada (NSERC).

■ REFERENCES

(1) Wright, P. E.; Dyson, H. Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999**, *293*, 321−331.
(2) Dunker, A.; Lawson, J.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; et al. Intrinsically Disordered Protein. *J. Mol. Graph. Modell.* **2001**, *19*, 26−59.
(3) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(4) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. Function and Structure of Inherently Disordered Proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756−764.
(5) Gsponer, J.; Madan Babu, M. The Rules of Disorder or Why Disorder Rules. *Prog. Biophys. Mol. Biol.* **2009**, *99*, 94−103.
(6) Baker, C. M.; Best, R. B. Insights into the Binding of Intrinsically Disordered Proteins from Molecular Dynamics Simulation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 182−198.
(7) Choi, U. B.; McCann, J. J.; Weninger, K. R.; Bowen, M. E. Beyond the Random Coil: Stochastic Conformational Switching in Intrinsically Disordered Proteins. *Structure* **2011**, *19*, 566−576.
(8) Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. Protein Disorder and the Evolution of Molecular Recognition: Theory, Predictions and Observations. *Pac. Symp. Biocomput.* **1998**, 473−484.
(9) Gunasekaran, K.; Tsai, C.-J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended Disordered Proteins: Targeting Function with Less Scaffold. *Trends Biochem. Sci.* **2003**, *28*, 81−85.
(10) Uversky, V. N. Unusual Biophysics of Intrinsically Disordered Proteins. *Biochim. Biophys. Acta—Proteins Proteom.* **2013**, *1834*, 932−951.
(11) Cino, E. A.; Choy, W.-Y.; Karttunen, M. Conformational Biases of Linear Motifs. *J. Phys. Chem. B* **2013**, *117*, 15943−15957.
(12) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. Intrinsic Disorder and Protein Function. *Biochemistry (Mosc.)* **2002**, *41*, 6573−6582.
(13) Uversky, V. N.; Dunker, A. K. Understanding Protein Non-Folding. *Biochim. Biophys. Acta—Proteins Proteom.* **2010**, *1804*, 1231−1264.
(14) Iakoucheva, L. M.; Brown, C. J.; Lawson, J.; Obradovic, Z.; Dunker, A. Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573−584.
(15) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.
(16) Vacic, V.; Iakoucheva, L. M. Disease Mutations in Disordered Regions—Exception to the Rule? *Mol. Biosyst.* **2012**, *8*, 27−32.
(17) Dunker, K. A.; Obradovic, Z.; Keith, A.; Obradovic, D. Z.; Romero, P.; Garner, E. C.; Brown, C. J. Intrinsic Protein Disorder in Complete Genomes. *Genome. Inform. Ser. Workshop Genome. Inform.* **2000**, *11*, 161−171.
(18) Pancsa, R.; Tompa, P. Structural Disorder in Eukaryotes. *PLoS One* **2012**, *7*, e34687.
(19) Hilser, V. J.; Thompson, E. B. Intrinsic Disorder As a Mechanism to Optimize Allosteric Coupling in Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8311−8315.
(20) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868−8873.
(21) Huang, Y.; Liu, Z. Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding−Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *J. Mol. Biol.* **2009**, *393*, 1143−1159.
(22) Chen, J. Towards the Physical Basis of How Intrinsic Disorder Mediates Protein Function. *Arch. Biochem. Biophys.* **2012**, *524*, 123−131.
(23) Schuler, B.; Hofmann, H. Single-Molecule Spectroscopy of Protein Folding Dynamics—Expanding Scope and Timescales. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36−47.
(24) Bernadó, P.; Svergun, D. I. Analysis of Intrinsically Disordered Proteins by Small-Angle X-ray Scattering. *Methods Mol. Biol.* **2012**, *896*, 107−122.
(25) Voelz, V. A.; Jager, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; et al. Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc.* **2012**, *134*, 12565−12577.
(26) Cino, E. A.; Killoran, R. C.; Karttunen, M.; Choy, W.-Y. Binding of Disordered Proteins to a Protein Hub. *Sci. Rep.* **2013**, *3*, 1−8.
(27) Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95*, L07−L09.
(28) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force Field Bias in Protein Folding Simulations. *Biophys. J.* **2009**, *96*, 3772−3780.
(29) Mittal, J.; Best, R. B. Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. *Biophys. J.* **2010**, *99*, L26−L28.
(30) Patapati, K. K.; Glykos, N. M. Three Force Fields' Views of the 3_{10} Helix. *Biophys. J.* **2011**, *101*, 1766−1771.
(31) Best, R. B.; Mittal, J. Free-Energy Landscape of the GB1 Hairpin in All-Atom Explicit Solvent Simulations with Different Force Fields: Similarities and Differences. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1318−1328.
(32) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(33) Best, R. B.; Mittal, J. Balance between α and β Structures in Ab Initio Protein Folding. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(34) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(35) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.
(36) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(37) Cino, E. A.; Choy, W.-Y.; Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725−2740.

(38) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(39) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7*, e32131.
(40) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, In the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(41) van Erp, T. S.; Bolhuis, P. G. Elaborating Transition Interface Sampling Methods. *J. Comput. Phys.* **2005**, *205*, 157−181.
(42) Hummer, G. From Transition Paths to Transition States and Rate Coefficients. *J. Chem. Phys.* **2004**, *120*, 516−523.
(43) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(44) E, W.; Ren, W.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(45) Faccioli, P.; Sega, M.; Pederiva, F.; Orland, H. Dominant Pathways in Protein Folding. *Phys. Rev. Lett.* **2006**, *97*, 108101.
(46) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(47) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(48) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271*, 997−999.
(49) Sotomayor, M.; Schulten, K. Single-Molecule Experiments In Vitro and In Silico. *Science* **2007**, *316*, 1144−1148.
(50) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(51) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(52) Michel, J.; Cuchillo, R. The Impact of Small Molecule Binding on the Energy Landscape of the Intrinsically Disordered Protein C-Myc. *PLoS One* **2012**, *7*, e41070.
(53) Baftizadeh, F.; Pietrucci, F.; Biarnes, X.; Laio, A. Nucleation Process of a Fibril Precursor in the C-Terminal Segment of Amyloid-β. *Phys. Rev. Lett.* **2013**, *110*, 168103.
(54) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6817−6822.
(55) Bian, Y.; Tan, C.; Wang, J.; Sheng, Y.; Zhang, J.; Wang, W. Atomistic Picture for the Folding Pathway of a Hybrid-1 Type Human Telomeric DNA G-Quadruplex. *PLoS Comput. Biol.* **2014**, *10*, e1003562.
(56) Itoh, K.; Chiba, T.; Takahashi, S.; Ishii, T.; Igarashi, K.; Katoh, Y.; Oyake, T.; Hayashi, N.; Satoh, K.; Hatayama, I.; Yamamoto, M.; Nabeshima, Y. An Nrf2/Small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes through Antioxidant Response Elements. *Biochem. Biophys. Res. Commun.* **1997**, *236*, 313−322.
(57) Itoh, K.; Wakabayashi, N.; Katoh, Y.; Ishii, T.; Igarashi, K.; Engel, J. D.; Yamamoto, M. Keap1 Represses Nuclear Activation of Antioxidant Responsive Elements by Nrf2 through Binding to the Amino-Terminal Neh2 Domain. *Genes Dev.* **1999**, *13*, 76−86.
(58) Tong, K. I.; Katoh, Y.; Kusunoki, H.; Itoh, K.; Tanaka, T.; Yamamoto, M. Keap1 Recruits Neh2 through Binding to ETGE and DLG Motifs: Characterization of the Two-Site Molecular Recognition Model. *Mol. Cell. Biol.* **2006**, *26*, 2887−2900.
(59) Lo, S.-C.; Li, X.; Henzl, M. T.; Beamer, L. J.; Hannink, M. Structure of the Keap1:Nrf2 Interface Provides Mechanistic Insight into Nrf2 Signaling. *EMBO J.* **2006**, *25*, 3605−3617.
(60) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(61) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry, and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(62) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting Biomolecular Flexibility with Metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148−154.
(63) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826−843.
(64) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium Free Energies from Nonequilibrium Metadynamics. *Phys. Rev. Lett.* **2006**, *96*, 090601.
(65) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(66) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714−6721.
(67) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(68) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197−2201.
(69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(70) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(71) Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. Spontaneous Formation of KCl Aggregates in Biomolecular Simulations: A Force Field Issue? *J. Chem. Theory Comput.* **2007**, *3*, 1851−1859.
(72) Chen, A. A.; Pappu, R. V. Parameters of Monovalent Ions in the AMBER-99 Forcefield: Assessment of Inaccuracies and Proposed Improvements. *J. Phys. Chem. B* **2007**, *111*, 11884−11887.
(73) Joung, I. S.; Cheatham, T. E. I. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(74) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(75) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **2014**, *114*, 779−814.
(76) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(77) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(78) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(79) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(80) Wishart, D. S.; Sykes, B. D.; Richards, F. M. The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure Through NMR Spectroscopy. *Biochemistry (Mosc.)* **1992**, *31*, 1647−1651.
(81) Wishart, D.; Sykes, B. The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data. *J. Biomol. NMR* **1994**, *4*, 171−180.
(82) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and Accurate Calculation of Protein ^{1}H, ^{13}C and ^{15}N Chemical Shifts. *J. Biomol. NMR* **2003**, *26*, 215−240.

(83) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Patternern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(84) Joosten, R. P.; te Beek, T. A. H.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W. W.; Schneider, R.; Sander, C.; Vriend, G. A Series of PDB Related Databases for Everyday Needs. *Nucleic Acids Res.* **2011**, *39*, D411−D419.
(85) Andersen, C. A.; Palmer, A. G.; Brunak, S.; Rost, B. Continuum Secondary Structure Captures Protein Flexibility. *Structure* **2002**, *10*, 175−184.
(86) Carter, P.; Andersen, C. A. F.; Rost, B. DSSPcont: Continuous Secondary Structure Assignments for Proteins. *Nucleic Acids Res.* **2003**, *31*, 3293−3295.
(87) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(88) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(89) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.
(90) Sibanda, B. L.; Thornton, J. M. β-Hairpin Families in Globular Proteins. *Nature* **1985**, *316*, 170−174.
(91) Cino, E. A.; Wong-ekkabut, J.; Karttunen, M.; Choy, W.-Y. Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response. *PLoS One* **2011**, *6*, e27371.
(92) Doshi, U.; Hamelberg, D. Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent. *J. Phys. Chem. Lett.* **2014**, *5*, 1217−1224.