JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

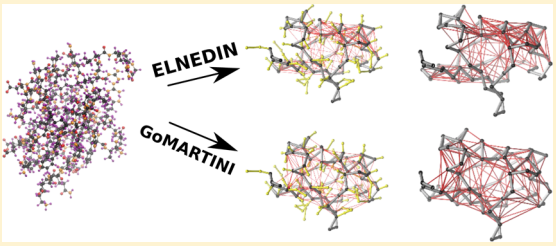
Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins

Adolfo B. Poma,* Marek Cieplak,* and Panagiotis E. Theodorakis*

Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Supporting Information

ABSTRACT: The application of coarse-grained (CG) models in biology is essential to access large length and time scales required for the description of many biological processes. The ELNEDIN protein model is based on the well-known MARTINI CG force-field and incorporates additionally harmonic bonds of a certain spring constant within a defined cutoff distance between pairs of residues, in order to preserve the native structure of the protein. In this case, the use of unbreakable harmonic bonds hinders the study of unfolding and folding processes. To overcome this barrier we have replaced the harmonic bonds with Lennard−Jones interactions based on the contact map of the native protein structure as is done in Gō-like models. This model exhibits very good agreement with all-atom simulations and the ELNEDIN. Moreover, it can capture the structural motion linked to particular catalytic activity in the Man5B protein, in agreement with all-atom simulations. In addition, our model is based on the van der Waals radii, instead of a cutoff distance, which results in a smaller contact map. In conclusion, we anticipate that our model will provide further possibilities for studying biological systems based on the MARTINI CG force-field by using advanced-sampling methods, such as parallel tempering and metadynamics.



1. INTRODUCTION

Computer simulations of proteins have contributed significantly to a better understanding of their structure and function by employing models that span a range of time and spatial scales. However, a higher spatial resolution of a model comes at the expense of higher computational cost, in this way restricting significantly the time scales accessible to simulation. To overcome this barrier, coarse-grained (CG) models for proteins have been developed in order to enable larger time and length scales in a simulation. In particular, CG models attempt to reduce the degrees of freedom of all-atom (AA) models by substituting a group of atoms with a single interaction site without loss of significant information that lies originally within the AA modeling of a system. A well-established CG model in the literature is the MARTINI force-field, which provides well-parametrized and transferable interaction potentials for proteins, as well as for a number of different molecules, for example, polymers, lipids, DNA, and others.[1−6] Advanced computational techniques that aim to couple AA and CG models, such as multiple resolution schemes for water molecules have employed this CG model due to its accuracy in reproducing atomistic results.[7] The success of MARTINI force-field is not only due to its capabilities to reformulate the AA force-field for a number of different molecules but also due to the intrinsic speed-up compared to AA simulations. For example, a time-scale speed-up factor of 4 has been estimated based on the diffusion dynamics of MARTINI water molecules compared to the atomistic single point charge model (SPC) water[8] or the case of peptide aggregation-rate of lipids into a bilayer.[9] Moreover, one may expect the whole dynamics of this CG model to become faster, due to the smaller number of particles in the CG model and the use of "softer" (CG) potentials than in the case of AA models. This results eventually in a quicker exploration of the phase space of the system and the possibility of using larger time steps in simulations for the integration of the equations of motion. Obviously, it is very difficult to quantify the speed-up in terms of a single conversion factor and generalize it for other systems, due to the multitude of length and energy scales involved in biological systems.[7,10]

In the case of proteins, the elastic network (EN) approach has been incorporated into the MARTINI force-field, in order to preserve the native structure of the protein.[11] To this end, the latest implementation of the EN strategy in the MARTINI force-field for proteins is the so-called ELNEDIN protein model.[12] In particular, ELNEDIN preserves the structure of a protein by adding harmonic bonds with a certain spring constant between C_α-atoms based on a cutoff criterion. However, the use of harmonic bonds prohibits studies of systems involving protein unfolding, which may restrict the possible applications of the MARTINI force-field for biological systems involving proteins. Moreover, including the harmonic bonds for nearest neighbor residues may increase significantly the number of interactions in the model (see Figure 1). In addition, there is no physically motivated criterion that

Received: October 7, 2016
Published: February 14, 2017

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

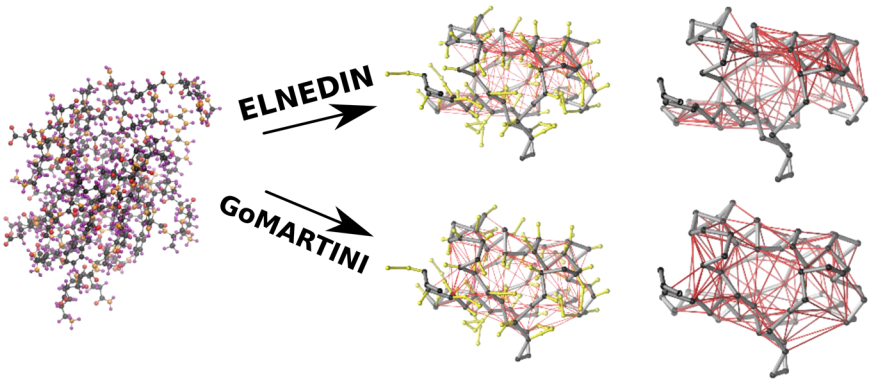


Figure 1. Snapshots showing the AA representation of the protein 1UBQ (left panel) and its CG representation in the ELNEDIN and GoMARTINI models as indicated. Water molecules are not shown for the sake of clarity. In the CG representations the rightmost column illustrates only the protein backbone and the network of bonds that preserve the structure of the protein for the sake of clarity, while the middle snapshots include also the side chains (yellow color). The red lines indicate the harmonic bonds and the LJ bonds in the ELNEDIN and in the GoMARTINI models, respectively. One can observe the broader distribution of contacts based on the contact map in the GoMARTINI approach. Moreover, we have 150 LJ-bonds in the case of GoMARTINI, while in the case of ELNEDIN, there are 272 harmonic bonds. The smaller number of bonds in the case of GoMARTINI can be also recognized visually. Snapshots have been produced with the VMD software.[27]

identifies the important contacts in the native structure of the protein.

Here, we have substituted the EN approach in the ELNEDIN protein model with Gō-like contact maps, which are mainly based on atomic overlaps.[13−18] In the Gō-like approach, some CG beads located at the positions of the C_α-atoms interact with a potential, which is often a Lennard−Jones (LJ). The selection of the C_α-atom pairs is based on the contact map, which is constructed from the native structure of the protein. In the MARTINI force-field, interactions are generally taken between groups of CG particles divided as polar (P), intermediately polar (N), apolar (C), and charged (Q). The first three types of interactions are represented by LJ potentials, whereas the last one is described by a Coulomb potential. The Gō-like interactions here are implemented as an additional set of LJ pairs. The main advantage of using such an approach is the ability to break the bonds based on the native contacts, as the harmonic bonds are replaced with LJ interactions. In this way, the model may allow for efficient sampling of unfolded states based on a statistical analysis of independent trajectories or advanced-sampling simulation protocols,[19] such as replica exchange,[20] metadynamics,[21] etc. Moreover, the Gō-like approach has been very successful in studying stretching and folding of proteins.[19,22] In the past, a few attempts have been carried out to explore large conformation changes in proteins based on the MARTINI force-field, for example, the study of the conformational flexibility of leucine binding protein and the gating mechanism of trans-membrane helices.[23,24] In this case, each single protein component is based on the ELNEDIN approach. As a result, no change at the level of the native structure for each single protein domain was possible, and only transitions due to the rearrangement of domains were allowed to take place. To this end, large conformations transitions occurring in protein folding and stretching cannot be studied by the ELNEDIN model due to the presence of harmonic bonds between amino acids in the EN, while further modifications of the structure-based patterns could provide possibilities of improving the sampling of unfolded states in a Gō-like approach.[25] Moreover, the additional LJ bonds are not added simply based on a cutoff criterion as in the EN approach, but according to the contact map based on the size of heavy atoms. As a result, the number of Gō bonds are almost half of what is required in the ELNEDIN model leading to a smaller number of interactions in the CG model. In addition, there is a number of different approaches to the contact map based on structural and chemical criteria, which could offer further improvement to the identification of the important contacts to be included in the model, as has been shown recently for a specific contact map.[13,26]

This paper presents the implementation of two contact maps in the MARTINI force-field in place of the EN approach. We have also explored the possibility of replacing the harmonic bonds in the EN approach with Gō-type LJ interactions. Moreover, we compare our simulation method with simulations based on ELNEDIN and AA modeling for several proteins, while discussing advantages and disadvantages of the method, as well as future directions. We anticipate that our work will provide further opportunities to tackle complex protein systems and processes by using the MARTINI force-field; in particular, in contexts involving large conformational transformations, such as during stretching, folding, and thermal unfolding.

2. METHODS

We used three different simulation methods in our investigation on the following proteins with PDB id: 1AOH (cohesin), 1TIT (titin), 1UBQ (ubiquitin), and 3W0K (Man5B). The sequential lengths of these proteins are 143, 89, 76, and 330 residues, respectively. The first three of these proteins have been often used in stretching studies,[28−31] while Man5B exhibits enzymatic activity that can be understood in terms of the equilibrium fluctuation.[18,31] The first method is the AA simulations, the second is based on the ELNEDIN protein model, and the third method is our approach involving two different choices of the contact map. To determine the contact map one considers each residue as a cluster of spheres. In this case, the radii of the spheres are equal to the van der Waals radii enhanced by a factor of about 25%. The first contact map takes into account the overlap of such spheres (OV),[14,32] while the second combines the OV with a variant of the contact of structure units (CSU).[33] The CSU approach takes into account the chemical properties of the atoms in contact. We denote the second contact map as OV+rCSU,[13] where the basic criterion in the rCSU is that two residues form a contact when the number of attractive atomic contacts between residues overcome the

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

repulsive ones (see Supporting Information (SI) for details on the OV and rCSU contact maps). We have also explored the possibility of using EN "contacts" as implemented in ELNEDIN and substituting the harmonic bonds by LJ interactions. Henceforth, we will refer to our approach based on the first two contact maps simply as "GoMARTINI" and indicate clearly the different contact maps used here as OV and OV+rCSU, respectively. In the case of Man5B, we compare our results on the principal component analysis (PCA) of the single-site positional fluctuation with AA-simulation results from the literature.[31] All simulations were carried out with the 4.6.5 version of the GROMACS package.[34−37] In the following, we provide details on the studied systems and the simulation protocols used.

All-Atom Simulations. Before starting our simulations, we checked the PDB protein structures for consistency and missing residues. Indeed, we found fully missing residues of the chain in the case of the Man5B protein in the PDB structure file, which we handled by using the MODELER software.[38−40] Other minor issues, regarding only partially missing side-chain atoms, were found in 1AOH, which have been already fixed in a previous study.[41] After these checks, we proceeded with the AA simulations following customary protocols in this field. In particular, we used the CHARMM27[42] force-field and the TIP3P water model[43] to simulate for each case a single protein in water. Initially, we solvated each protein in a cubic simulation box with a distance of 1.2 nm between the box edge and the protein. This distance is large enough to guarantee the minimum image convention and more than sufficient for any cutoff scheme used in our simulations. In the next step, we check the amount of total charge on the protein and we subsequently add the necessary ions in order to neutralize the charge of the protein. Later, we performed minimization of the systems by carefully monitoring the potential energy and the maximum force in the system. This minimization procedure guarantees that we have a reasonable initial configuration for equilibration and production runs. The procedure toward system equilibrium was implemented in two steps. In the first step, we ran Molecular Dynamics (MD) simulations in the canonical ensemble (*NVT*) using a velocity rescaling thermostat[44] for 200 ps to stabilize the temperature of the system at 300 K. In the second step, we ran simulations for 400 ps in the isothermal−isobaric ensemble (*NPT*) to stabilize the pressure of the system at 1 bar using the Parrinello-Raman barostat[45] monitoring, also, the density. The same thermodynamic conditions were chosen for the ELNEDIN and GoMARTINI models. Electrostatic interactions in the AA MD simulations were included using the Particle Mesh Ewald (PME) method,[46,47] whereas in the ELNEDIN and GoMARTINI methods, it was computed by using a reaction field method[48,49] with a relative dielectric constant (ϵ = 15). After these initial equilibration steps we carried out 100 ns production runs for each case, monitoring carefully all the necessary thermodynamic parameters of the systems. Our results are based on the analysis of these production runs.

ELNEDIN Simulations. The ELNEDIN protein model and version 2.2 of the MARTINI force-field were used to simulate the four proteins.[4,12] In the case of ELNEDIN, the backbone beads (BB) are placed at the positions of the C_α-atoms, instead of the center of mass of the N, C_α, C_β, and O atoms of the atomistic backbone that was originally implemented in the MARTINI force-field for proteins.[3] In the ELNEDIN model, the side chains are present in the model as additional CG beads according to the parametrization of the MARTINI force-field. The same definition of CG side chains is followed in the case of the GoMARTINI model. Due to this change, small modifications on the bonded interactions and the structural mapping of aromatic residues only were introduced in the ELNEDIN model of the MARTINI force-field, while the nonbonded interactions were not modified.[12] The EN scaffolding component in the ELNEDIN model was implemented by applying harmonic bonds of a certain strength between CG BB located at the C_α-atoms being within a cutoff distance. In the ELNEDIN model, both the spring constant of the harmonic bond potential and the cutoff are optimized being the same for each pair of CG BB.[12] Values for the spring constant between 500 and 1000 kJ mol^{-1} nm^{-2} and cutoff distances between 0.8 and 1.0 nm provide adequate quantitative agreement with AA simulations.[12] In our case, we have used the benchmark values of 0.9 nm for the cutoff and 500 kJ mol^{-1} nm^{-2} for the harmonic constant.

The equations of motion for each bead were integrated by using the velocity Verlet algorithm with a fixed integration time step during the simulation. We assume the same time scale as in the AA simulations in our study, since the conversion factor which reflects the speed up for proteins has not yet been determined. While the time step in AA simulations is of the order of 1 fs, CG simulations enable the use of a larger time step. The simulation box for each protein reached the following dimensions: L_x = 7.3 nm, L_y = 7.3 nm, and L_z = 7.9 nm. The periodic boundary conditions were used to avoid the problem of the finite size effects. Initially, a trivial minimization was performed in vacuum for 100 ps with a time step of 1 fs and position restraints (1000 kJ mol^{-1} nm^{-2}) applied to all C_α atoms. Later, each protein was solvated with a minimum of 1.2 nm between any bead and the box edge and additional energy minimization with the same parameters took place. Our systems were further relaxed by using 2 ns long MD simulations with a time step of 20 fs in the presence of position restraints. At the end, we ran 200 ns MD runs without the restraints. For our MD simulations, we used here also the modified velocity rescaling thermostat[44] acting on each single CG bead and the Parrinello−Rahman barostat[45] in the same thermodynamic conditions as in the AA simulations. Our analysis is based on the last 100 ns of the simulated trajectories consisting of 2000 samples.[12]

GoMARTINI Simulations. In the GoMARTINI model we have replaced the harmonic bonds with LJ potential interactions between CG BB according to the contact map of the native protein structure as is done in Gō models.[16,22] The LJ potential reads

$$U_{\mathrm{LJ}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \quad (1)$$

There are different flavors of contact maps[26,50] and arising differences can be found in the total number of contacts. In the present study we have chosen at first the simplest approach based on the atomic overlap criterion (OV).[16,22] In this case, the parameter σ_{ij} of the LJ potential is determined by the following relationship valid for LJ potential, $\sigma_{ij} = d/2^{1/6}$, with d being the C_α−C_α distance between a pair of residues that form a contact. Because of the presence of the MARTINI force-field, nonbonded interactions are also described by the LJ potentials, but, in this case, the parameter ε_{ij} varies in the range of 2.0−5.6 kj mol^{-1} according to the MARTINI CG bead groups (P, N,

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

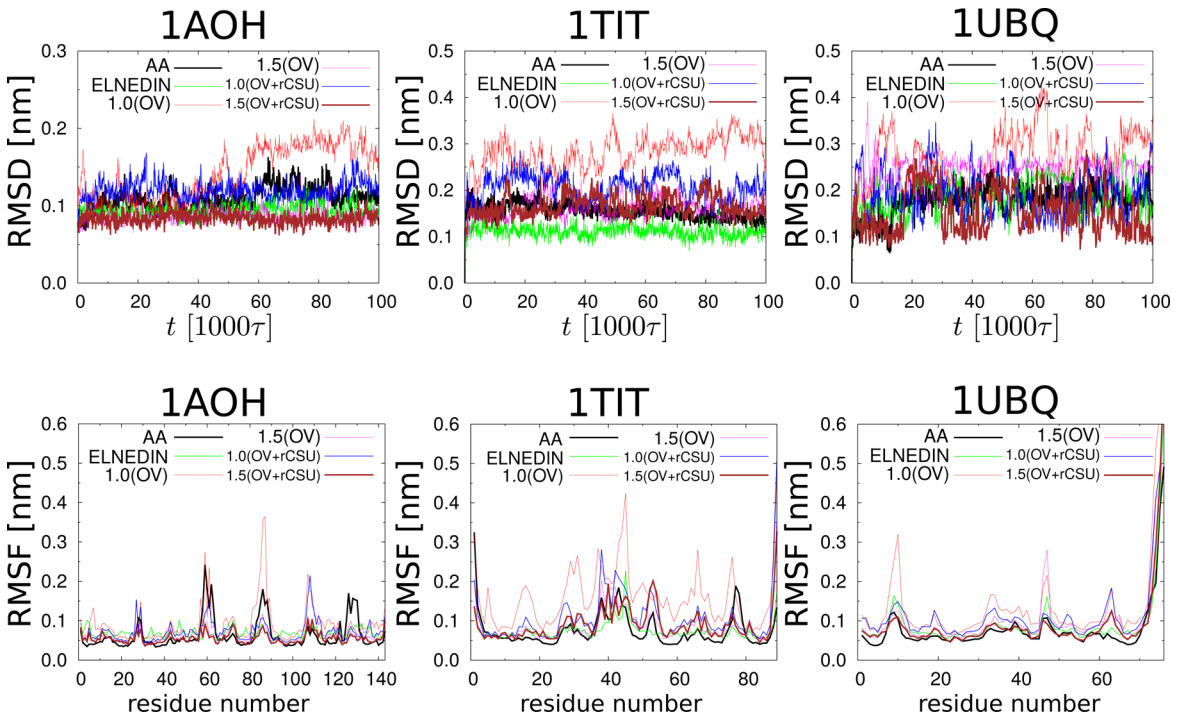


Figure 2. RMSD and RMSF based on the GoMARTINI model in the case of three proteins (1AOH, 1TIT, and 1UBQ) for two values of λ = 1.0 and 1.50, and two different contact maps (OV and OV+rCSU), as indicated. The comparison with the ELNEDIN and AA models is also illustrated in each graph. The corresponding time-scale in the AA results is given by τ = 1 ps. Comparison with ELNEDIN and GoMARTINI simulations were done by assuming the same time scale.

and C). In Gō-like models this parameter is usually set to a larger value, which typically represents the strength of hydrogen bonds (HB) and other contributions, such as ionic bridges and other kinds of specific interactions (e.g., hydrophobic).[51] In the GoMARTINI, the parameter ε_{ij} is expressed in units of ε, where ε = 6.276 kj mol^{-1}. This value has been estimated previously from AA simulations,[18] and it corresponds to the typical energy scale of hydrogen bonds in proteins. Hence, the parameter λ in the native contact energy, $\varepsilon_{ij} = \lambda\ \varepsilon$, is a tunable parameter of the GoMARTINI model. The parameter λ is simply a number indicating the strength of the LJ potential, and an optimum value for ε_{ij} can be found in order to match quantitatively the GoMARTINI to the AA model. Thus, the LJ bonds behave in the same way that the harmonic bonds act in the EN approach implemented in the ELNEDIN,[11] except that they are breakable. In this study, we have also implemented in the GoMARTINI model the OV+rCSU contact map, which includes additional contacts (about 8−10% of the total number of contacts obtained from OV) and has shown a better agreement with experimental data.[13] Similarly to that of the ELNEDIN, the same simulation protocol was followed here for the GoMARTINI simulations.

Pulling Simulations. We carried out pulling simulations based on the GoMARTINI model, as is typically done using Gō-like models.[19] In our case, however, pulling of the protein took place in the presence of explicit solvent. In the case of the standard MARTINI water, four atomistic water molecules are mapped onto a CG water bead. In pulling simulations, the protein is pulled along the end-to-end vector connecting the C_α-atoms from N- and C-terminus and the reaction coordinate is the displacement of the pulling spring. Moreover, additional beads have been attached to those C_α-atoms with the spring constant being 37.6 kj mol^{-1} nm^{-2}, which is a typical value of the AFM cantilever stiffness in protein stretching studies. Each system was pulled over the course of 300 000 ps with a velocity of 10^{-5} nm/ps. Although this value is still far from the experimental values of cantilever velocities ($\sim 10^{-9}$ nm/ps[28−30]), it represents a significant computational improvement to access more realistic time-scales compared to AA simulations,[52] where in the latter case a large speed ($\sim 10^{-2}$ nm/ps) is typically used to compromise on the computational demand even in high-performance computing facilities. In experiments, multiple proteins are linked sequentially, and one can observe a number of corresponding peaks, which signal the full unfolding of one protein module. Because of the space resolution, intermediate unfolding states are not detected in AFM experiments. However, in the case of the Gō-like model usually one can access these intermediate states with a better resolution and assign to each of them a force peak. The largest of these force peaks, F_{max}, defines the characteristic unfolding force.

Folding Simulations. We studied the folding process of two small peptides well documented in the literature, an α-helix comprising residues 70−83 of the protein HPr (PDB: 1HDN with 85 residues in total) and a β-hairpin (residues 41−56 of the protein G with PDB 1GB1 and 56 residues in total).[53] Folding simulations required the preparation of initial configurations without the presence of any native contacts (unfolded structures). Native contacts are present in the structure when the actual distance between two BB is smaller than $1.5\sigma_{ij}$ or $1.3d$ where d is the distance between C_α atoms in the native structure.[19] The coordinates of unfolded structures were obtained by heating up the protein CG structure at 500 K without water. In this way, we could produce a number of independent unfolded structures, which served as intitial configurations for 50 independent simulation trajectories of

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

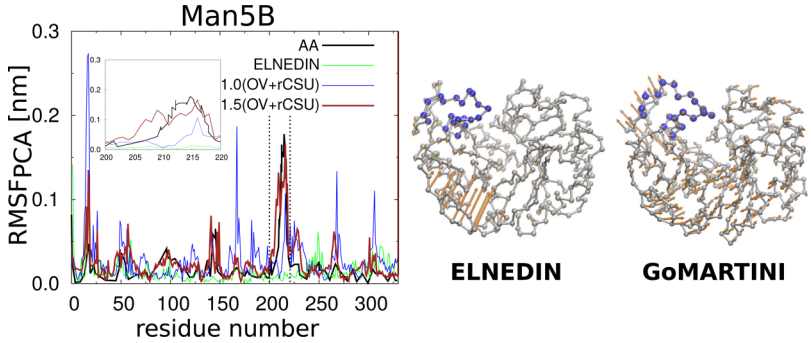


Figure 3. Principal component analysis of the RMSF for the C_α-atoms (AA) or the CG BB (ELNEDIN and GoMARTINI) of the Man5B. We present results obtained from AA, ELNEDIN, and GoMARTINI models based on the OV+rCSU contact map, as indicated. The AA data were taken from Figure 4 in ref 31 by using the g3data software.[57] The inset shows the fluctuation from the catalytic region between the 200 and 220 amino acid residues in Man5B. This figure illustrates the excellent agreement of the GoMARTINI model based on the OV+rCSU contact map with AA data for the choice $\lambda = 1.5$. The principal component is illustrated for the case of ELNEDIN and GoMARTINI for each BB, where BB in blue color are part of the catalytic pocket (200−220). In the GoMARTINI case, fluctuations along the principal component for the catalytic pocket are in line with the expected behavior observed in all-atom simulations.[31] Snapshots in this figure have been produced with the VMD[27] and the PRODY software.[58,59] Water molecules are not shown for the sake of clarity.

100 ns at 300 K in the presence of water following the protocols described above for the GoMARTINI simulations. Our results are based on the analysis of these trajectories.

Principal Component Analysis. MD trajectories involve high dimensional data, and therefore, it is difficult to identify data patterns that correspond to correlated motions important for biological processes. The PCA is a simple method that succeeds in emphasizing the essential dynamics of a protein by rendering data easy to explore and visualize.[18,54−56] This is achieved by reducing the number of dimensions, without much loss of information. The PCA involves the calculation of the covariance matrix and its eigenvectors and eigenvalues. The eigenvector with the largest eigenvalue is the "principal component" of the data set, and usually eigenvectors are sorted according to the eigenvalues from highest to lowest corresponding, also, to the order of significance. If we ignore components, then we may lose some information on the original data set, but the loss is small given that the corresponding eigenvalue is also small. The final step of the PCA involves the choice of the relevant eigenvectors put in a matrix form and multiplication with the original data in order to obtain the new data set in terms of the vectors that we have chosen. In our case, the PCA was used for the single-site fluctuations of the backbone atoms/beads of the Man5B protein, which exhibits enzymatic activity that manifests itself in a characteristic motion of a part of the protein between residues 200 and 220, which is very nicely captured in AA simulations in the literature.[31] In our study, we compared the ELNEDIN and GoMARTINI models with the latter PCA data from the AA simulations.[31]

3. RESULTS AND DISCUSSION

Root Mean Square Deviation (RMSD) of CG BB and C_α-Atoms. In our study, we have calculated the RMSD over the course of the last 100 ns of our 200 ns trajectories for three proteins (1AOH, 1TIT, and 1UBQ). Our results based on the GoMARTINI model indicate that the proteins were stable during the simulations with deviations below 0.2 nm (see Figure 2), in a very good agreement with the AA simulations. The GoMARTINI performs generally well for the three different proteins for both the OV and the OV+rCSU contact maps and is based on a smaller number of contacts. On the contrary, ELNEDIN seems to contribute to an increased backbone stiffness of the protein for the choice of spring constant and cutoff used in this study. Of course, the values of the spring constant and the cutoff in the ELNEDIN model can be tuned in order to fit the AA simulations better.[12] However, we do not aim here at a direct comparison between the GoMARTINI and ELNEDIN models. Moreover, the use of the EN-approach for the contact map, in which we substituted the harmonic bonds with LJ interactions, had delivered results similar to the ELNEDIN approach with the magnitude of fluctuations depending on the choice of λ indicating that the choice of contact map is a very important element in order to reproduce the AA behavior of proteins (see Supporting Information). In the GoMARTINI the tunable parameter of the model is the factor λ in the LJ potential, which is used for the native contacts for a given choice of the contact map.[13] By scanning a range of values for λ between 0.25 and 2.00, we found that values of λ around 1.5 give very good results in agreement with AA simulations, for both the OV and OV +rCSU contact maps, with the latter exhibiting a slightly better performance than the OV contact map. Here, it seems that the optimum value of λ is the same for all proteins studied with GoMARTINI. Note that a larger value of ε_{ij}, 50% stronger than typical HB strength (ϵ), is needed to stabilize the native structure. One may suggest that this effect may be due to a larger number of CG beads (2−5 per residue) in the MARTINI force-field in comparison with Gō-like models, where a residue is merely represented by the C_α atom. In the latter model ε_{ij} corresponds to the strength of HB and the reduced temperature of $T = 0.3\varepsilon/k_B$ with $\varepsilon_{ij} = \varepsilon$ being generally used.[13,16,19,22] This also indicates the strong character of the LJ interactions in the GoMARTINI model and its correspondence to the EN approach of the ELNEDIN in the limit of harmonic approximation of the LJ potential.[11]

Root Mean Square Fluctuations (RMSF) of CG BB and C_α-Atoms. The RMSF for the CG BB in the GoMARTINI model is able to capture the characteristic fluctuations at most of the residues showing very good agreement with the ELNEDIN and AA simulations (see Figure 2), where for the RMSF of AA simulations the positions of the C_α-atoms were considered for our analysis. Moreover, our results for the RMSF support a value of ε_{ij} around 1.5ε for the range of discrete

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

values used in this study. In the RMSF analysis, it becomes more apparent that the OV+rCSU contact map may be a better choice than the OV contact map, again indicating the importance of choosing the contacts properly.

Principal Component Analysis. AA simulations have illustrated the key PCA mode involved in the opening and closing motion of the Man5B catalytic pocket (residues between 200 and 220 along the protein chain) capturing the expected motional amplitude of each amino acid in the PCA for Man5B with and without any substrate.[31] A Gō-like model has been employed to describe the same system. Although it does differentiate the strength of the collective motion in the presence of the substrates, the agreement with AA is not full.[18] Our results based on the GoMARTINI model (OV+rCSU contact map) are in excellent agreement with the AA simulations for the case without substrate and most importantly GoMARTINI is able to capture the opening and closing motion of the Man5B catalytic pocket, which exhibits the same amplitude in the PCA as in the AA simulations for the optimum choice of $\lambda = 1.5$ (see Figure 3). On the other hand, a choice of $\lambda = 1.0$ leads to higher flexibility for parts of the protein, and, also, affects the height of the peak in the enzymatic pocket (residues 200−220 illustrated as inset of Figure 3). Moreover, we found that the use of OV+rCSU contact map yields better results than the OV contact map (see Supporting Information). The PCA analysis of Man5B based on the ELNEDIN protein model does not capture this characteristic motion of the enzymatic pocket. Moreover, a visual inspection of the fluctuations based on the PCA indicates the opening and closing of the catalytic pocket in the case of the GoMARTINI approach, in agreement with all-atom simulations.[31] On the contrary, ELNEDIN results in a high stiffness in most parts of the protein and large fluctuations in a certain part of the protein structure deviating considerably from the results based on all-atom simulations.[31] Hence, it seems that the information on the native contacts of the protein, which is the criterion for adding the LJ bonds in the GoMARTINI model, may be essential in the case of the Man5B for selecting the pairs of contacts, instead of adding harmonic bonds based only on a cutoff criterion.[12] The latter conclusion is further corroborated by the third contact map explored in this study based on the EN approach and LJ interactions. Indeed, the main conclusion of our study on different proteins leads to the significance of choosing "properly" the contact map,[31] while we have confirmed that the OV+rCSU delivers better results for all cases considered here, in agreement with a previous study, which explored various Gō-like models.[13] Furthermore, one should also take into account differences that may arise due to the different environment of the C_α-atoms in the AA models and the CG BB.

Pulling Simulations. We conducted pulling simulations by using the GoMARTINI model for the I27 domain of titin (1TIT) protein. Figure 4a shows typical stretching curves, in which two different contact maps (i.e., OV and OV+rCSU) give rise to a change in F_{max}. In particular, we obtain higher peaks in the case of the OV+rCSU contact map due to the additional contacts in comparison with the OV contact map. Moreover, the case with $\lambda = 1.5$ displays a subsequent small second peak in agreement with Gō-like simulations for this system.[19] Also, the nonbonded MARTINI interactions play a significant role, as they are actively contributing during the stretching process. To study this contribution we performed first an ELNEDIN simulation for 100 ns; then, all harmonic bonds from the EN

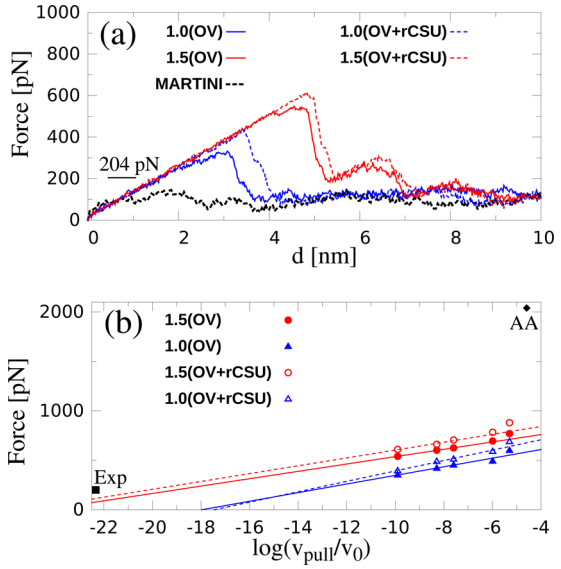


Figure 4. Pulling results with GoMARTINI for the I27 domain of titin and for two values of $\lambda = 1.0$ and 1.50 and two contact maps (OV and OV+rCSU), as indicated. Panel a illustrates stretching curves with a pulling velocity 10^{-5} nm/ps. MARTINI curves were computed by removing the harmonic bonds of the EN. Panel b shows F_{max} as a function of pulling velocity with $v_0 = 1$ nm/ps. The experimental (square point) and AA (diamond point) values of F_{max} shown in the panel b are taken from refs 29 and 60, respectively.

were removed and the pulling simulations were finally carried out. Figure 4a illustrates the background signal due to the MARTINI force field, which gives essentially no force peaks and its pattern has a threshold of about 100 pN. In Figure 4b we present the dependence of F_{max} on the pulling speed in the range of 10^{-5}−10^{-3} nm/ps, where GoMARTINI force-field provides a good estimate for F_{max}. For the sake of comparison, AA stretching for this protein gives a high value of $F_{max} \approx 2040$ pN, mainly due to the large pulling speed. AFM experiments are typically carried out at very slow speeds compared to AA. Although our pulling speeds are still larger than experimental ones, they represent a significant improvement with respect to AA models allowing for comparison with experiments. Indeed, an extrapolation of our data for each λ for the cases of OV and OV+rCSU contact maps shows that the fitting curves approach well the experimental value for $\lambda = 1.5$. Our conclusions are further corroborated by pulling simulations for the 1AOH and 1UBQ cases (see Supporting Information). Note that these results correspond to a set of well-resolved experimental protein structures. For example, we do not expect the same full agreement for very large proteins. In this case, the methodology presented will be limited for the case of defected protein structures with several missing side chains, as it relies on the determination of native contacts based on the positions of the heavy atoms. However, if the number of defected amino acids is not large, one can still rely on basic modeling to reconstruct them, otherwise this methodology will be unpractical.

Folding Simulations. GoMARTINI allows for the study of protein folding, whereas the ELNEDIN model is not suitable for folding studies because of the presence of the harmonic bonds required to maintain the native structure of the protein. Moreover, the MARTINI force-field alone cannot stabilize the structure of a protein without the presence of these harmonic bonds. Here, we have enhanced our discussion on systems undergoing large conformational changes by including the

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

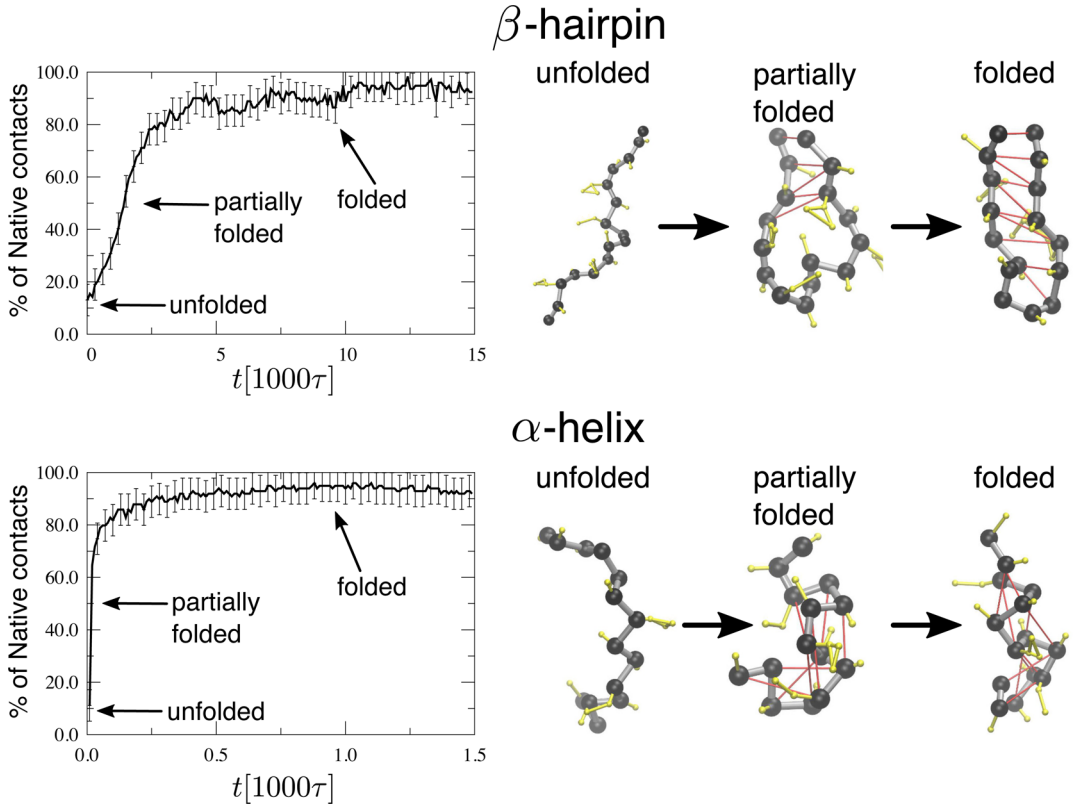


Figure 5. Folding of two small peptides, a β-hairpin and an α-helix (see section 2 for details). Plots show the percentage of native contacts present at a certain time during the folding process for each case, as indicated. A typical folding time for the α-helix is about 1 ns, while for the β-hairpin it is about 10 ns. Snapshots indicate examples of an unfolding configuration of BB at the beginning of a folding simulation, a partially folded structure, and a final folded (native) structure for each peptide. The scale of these snapshots is different. The BB and side chain beads are represented by black and yellow color, respectively. Native contacts during folding are described by solid red lines. Snapshots in this figure have been produced with the VMD software.[27] Water molecules are not shown in the snapshots for the sake of clarity.

folding of an α-helix and a β-hairpin (see section 2 for details).[53] A summary of our results and snapshot examples of unfolded, partially folded (50% of native contacts present in the structure), and folded (native) structures are presented in Figure 5. In the case of the α-helix, folding did occur for all independent trajectories, whereas in the case of the β-hairpin 4% of the trajectories did not lead to the native folded structure within the 100 ns time. In agreement with previous studies,[53] we find that folding of the α-helix takes place about 10 times faster than the folding of the β-hairpin. In particular, folding of the α-helix is about 1 ns, while in the case of the β-hairpin the folding time is about 10 ns (Figure 5). However, comparison of the folding times on absolute terms between our GoMARTINI simulations and experiments (folding times are of the order of microseconds) is not currently possible, because the time scale in simulations has been only roughly known.

4. CONCLUSIONS

The present work has underlined the possibility of using the protein contact map to substitute the harmonic bonds in the ELNEDIN protein model with LJ interactions as is done in Gō models, in this way enabling the study of unfolding processes or simply the consideration of unfolded protein states in advanced-sampling simulations by using the MARTINI force field. Moreover, our approach makes use of the contact map, which identifies the key pairs of contacts between residues required to preserve the native structure of the protein without any adjustable parameters. Our results are in very good agreement with the ELNEDIN protein model and the AA simulations, for which the protein has not undergone any large conformational changes. Furthermore, the GoMARTINI model is able to capture the key motion related to a certain catalytic activity in the case of the protein Man5B, in agreement with AA simulations. We also find that the use of OV+rCSU as a contact map constitutes the best choice from the contact maps considered in our study, while a choice of roughly $\varepsilon_{ij} = 1.5\varepsilon$ for the well depth of the LJ contact potential is well-suited to guarantee agreement with AA data. In addition, this value has been shown to approach the experimental values of F_{max} in stretching simulations for the I27 domain of titin at the experimental speed after extrapolation. Finally, the GoMARTINI model has enabled the study of protein folding confirming the 10-fold difference in folding time between the α-helix and the β-hairpin.[53] We anticipate that our study may offer new venues in the CG simulation arena of biological molecules undergoing large conformational changes based on the MARTINI force-field.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00986.

Further details on the contact maps, pulling results for 1AOH and 1UBQ proteins and results based on the ELNEDIN with LJ interactions and PCA for RSMF of the Man5B based on the OV contact map as obtained

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

from the GoMARTINI; additional tables and figures (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: poma@ifpan.edu.pl.
*E-mail: mc@ifpan.edu.pl.
*E-mail: panos@ifpan.edu.pl.

ORCID

Adolfo B. Poma: 0000-0002-8875-3220

Funding

This research has been supported by the National Science Centre, Poland, under Grant No. 2014/15/B/ST3/01905 and Grant No. 2015/19/P/ST3/03541 and the European Framework Programme VII NMP Grant 604530-2 (CellulosomePlus). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska−Curie Grant Agreement No. 665778. This research was supported in part by PLGrid Infrastructure.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Bartosz Różycki for useful discussions and Karol Wołek for providing the rCSU contacts.

REFERENCES

(1) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.
(2) Marrink, S.; Risselada, H.; Yefimov, S.; Tieleman, D.; de Vries, A. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(3) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(4) de Jong, D.; Singh, G.; Bennett, W.; Arnarez, C.; Wassenaar, T.; Schäfer, L.; Periole, X.; Tieleman, D.; Marrink, S. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.
(5) Herzog, F. A.; Braun, L.; Schoen, I.; Vogel, V. Improved Side Chain Dynamics in MARTINI Simulations of Protein-Lipid Interfaces. *J. Chem. Theory Comput.* **2016**, *12*, 2446.
(6) Lelimousin, M.; Limongelli, V.; Sansom, M. S. P. Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained Metadynamics Free Energy Calculations. *J. Am. Chem. Soc.* **2016**, *138*, 10611.
(7) Zavadlav, J.; Melo, M. N.; Marrink, S. J.; Praprotnik, M. Adaptive Resolution Simulation of an Atomistic Protein in MARTINI Water. *J. Chem. Phys.* **2014**, *140*, 054114.
(8) Fuhrmans, M.; Sanders, B. P.; Marrink, S.-J.; de Vries, A. H. Effects of Bundling on the Properties of the SPC Water Model. *Theor. Chem. Acc.* **2010**, *125*, 335−344.
(9) Thøgersen, L.; Schiøtt, B.; Vosegaard, T.; Nielsen, N. C.; Tajkhorshid, E. Peptide Aggregation and Pore Formation in a Lipid Bilayer: A Combined Coarse-Grained and All Atom Molecular Dynamics Study. *Biophys. J.* **2008**, *95*, 4337−4347.
(10) Zavadlav, J.; Melo, M. N.; Marrink, S. J.; Praprotnik, M. Adaptive Resolution Simulation of Polarizable Supramolecular Coarse-Grained Water Models. *J. Chem. Phys.* **2015**, *142*, 244118.
(11) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(12) Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A. Combining an Elastic Network with a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.
(13) Wołek, K.; Gómez-Sicilia, A.; Cieplak, M. Determination of Contact Maps in Proteins: A Combination of Structural and Chemical Approaches. *J. Chem. Phys.* **2015**, *143*, 243105.
(14) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* **1999**, *290*, 253−66.
(15) Hoang, T. X.; Cieplak, M. Sequencing of Folding Events in Go-Type Proteins. *J. Chem. Phys.* **2000**, *113*, 8319−8328.
(16) Sułkowska, J. I.; Cieplak, M. Selection of Optimal Variants of Gō-like Models of Proteins Through Studies of Stretching. *Biophys. J.* **2008**, *95*, 3174−3191.
(17) Sikora, M.; Sułkowska, J. I.; Cieplak, M. Mechanical Strength of 17 134 Model Proteins and Cysteine Slipknots. *PLoS Comput. Biol.* **2009**, *5*, 1−15.
(18) Poma, A. B.; Chwastyk, M.; Cieplak, M. Polysaccharide-Protein Complexes in a Coarse-Grained Model. *J. Phys. Chem. B* **2015**, *119*, 12028−12041.
(19) Cieplak, M.; Hoang, T.; Robbins, M. Folding and Stretching in a Go-like Model of Titin. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 114−124.
(20) Swendsen, R.; Wang, J. Replica Monte Carlo Simulation of Spin Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(21) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(22) Sułkowska, J. I.; Cieplak, M. Mechanical Stretching of Proteins−A Theoretical Survey of the Protein Data Bank. *J. Phys.: Condens. Matter* **2007**, *19*, 283201.
(23) Siuda, I.; Thøgersen, L. Conformational Flexibility of the Leucine Binding Protein Examined by Protein Domain Coarse-Grained Molecular Dynamics. *J. Mol. Model.* **2013**, *19*, 4931−4945.
(24) Wang, Z.; Liao, J.-L. Probing Structural Determinants of ATP-Binding Cassette Exporter Conformational Transition Using Coarse-Grained Molecular Dynamics. *J. Phys. Chem. B* **2015**, *119*, 1295−1301.
(25) 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.
(26) Noel, J.; Whitford, P.; Onuchic, J. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692−8702.
(27) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(28) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. Mechanical and Chemical Unfolding of a Single Protein: A Comparison. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3694−3699.
(29) Marszalek, P. E.; Lu, H.; Li, H.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. M. Mechanical Unfolding Intermediates in Titin Modules. *Nature* **1999**, *402*, 100−103.
(30) Valbuena, A.; Oroz, J.; Hervás, R.; Vera, A. M.; Rodríguez, D.; Menéndez, M.; Sułkowska, J. I.; Cieplak, M.; Carrión-Vázquez, M. On the Remarkable Mechanostability of Scaffoldins and the Mechanical Clamp Motif. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13791−6.
(31) Bernardi, R.; Cann, I.; Schulten, K. Molecular Dynamics Study of Enhanced Man5B Enzymatic Activity. *Biotechnol. Biofuels* **2014**, *7*, 83.
(32) Chwastyk, M.; Jaskólski, M.; Cieplak, M. The Volume of Cavities in Proteins and Virus Capsids. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1275−1286.
(33) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated Analysis of Interatomic Contacts in Proteins. *Bioinformatics* **1999**, *15*, 327−332.
(34) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374

(35) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(36) 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.
(37) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.; Smith, J.; Kasson, P.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(38) Marti-Renom, M.; Stuart, A.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.
(39) Sali, A.; Blundell, T. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(40) Fiser, A.; Do, R.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9*, 1753−1773.
(41) Chwastyk, M.; Poma, A. B.; Cieplak, M. Statistical Radii Associated with Amino Acids to Determine the Contact Map: Fixing the Structure of a Type I Cohesin Domain in the Clostridium Thermocellum Cellulosome. *Phys. Biol.* **2015**, *12*, 046002.
(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
(44) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(45) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182.
(46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(47) 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.
(48) Neumann, M. Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems. *Mol. Phys.* **1983**, *50*, 841−858.
(49) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(50) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(51) Sheu, S.-Y.; Yang, D.-Y.; Selzle, H.; Schlag, E. Energetics of Hydrogen Bonds in Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12683−12687.
(52) Lu, H.; Schulten, K. Steered Molecular Dynamics Simulations of Force-Induced Protein Domain Unfolding. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 453−463.
(53) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The Protein Folding 'Speed Limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(54) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.
(55) Yang, L. W.; Eyal, E.; Bahar, I.; Kitao, A. Principal Component Analysis of Native Ensemble of Biomolecular Structures (PCA_NEST): Insights Into Functional Dynamics. *Bioinformatics* **2009**, *25*, 606−614.
(56) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996**, *100*, 2567−2572.
(57) Bauer, B.; Reynolds, M. Recovering Data from Scanned Graphs: Performance of Frantz's g3data Software. *Behav. Res. Methods* **2008**, *40*, 858−868.
(58) Bakan, A.; Meireles, L.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575−1577.
(59) Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T.; Bahar, I. Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics. *Bioinformatics* **2014**, *30*, 2681−2683.
(60) Lui, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation. *Biophys. J.* **1998**, *75*, 662−671.

DOI: 10.1021/acs.jctc.6b00986
J. Chem. Theory Comput. 2017, 13, 1366−1374