Disorder–Order Transitions in Conformational Selection of a Peptide by Ebola Virus Nucleoprotein

Mark A. Olson*

ABSTRACT: This study presents parallel-tempering lattice Monte Carlo simulations based on the side-chain-only (SICHO) model for calculating the conformational landscape of a 28-residue intrinsically disordered peptide extracted from the Ebola virus protein VP35. The central issue is the applicability of the SICHO potential energy function and in general coarse-grained (CG) representations of intermediate resolution for modeling large-scale conformational heterogeneity that includes both folded and unstructured peptide states. Crystallographic data shows that the peptide folds in a 4\(_{10}\)-helix-turn-3\(_{10}\)-helix topology upon complex formation with the Ebola virus nucleoprotein, whereas in isolation, the peptide transitions to a disordered conformational ensemble as observed in circular dichroism experiments. The simulation reveals a potential of mean force that displays conformational diversity along the helix-forming reaction coordinate consistent with disorder–order transitions, yet unexpectedly the bound topology is poorly sampled, and a population shift to an unstructured state incurs a significant free-energy penalty. Applying an elastic network interpolation model suggests a hybrid binding mechanism through conformational selection of the 4\(_{10}\)-helix followed by an induced fit of the 3\(_{10}\)-helix. A comparison of the CG model with previously reported all-atom CHARMM-based simulations highlights a lattice-based approach that is computationally fast and with the correct parameterization yields good resolution to modeling conformational plasticity.

1. INTRODUCTION

The Ebola virus encodes seven structural proteins of which most occupy unique points in topological fold space. Given the proteome size and repertoire of globular domains, the interactome is thought to be largely multifunctional during the virus lifecycle. Elucidation of the landscape of viral protein interactions in terms of association networks is of fundamental importance for the discovery of effective therapeutics. A step forward is the recent X-ray crystallographic structure of a 28-residue peptide extracted from the Ebola virus protein VP35 in association with the Ebola virus nucleoprotein. The peptide folds in a 4\(_{10}\)-helix-turn-3\(_{10}\)-helix topology upon binding the nucleoprotein, while, in isolation, a disordered ensemble of states is observed in circular dichroism (CD) experiments.

The coupled folding and binding suggests that the peptide can be broadly characterized as an intrinsically disordered peptide (IDP). What is intriguing is when the peptide is added to a solution of 50% trifluoroethanol (TFE), CD experiments show a transition from a predominately unstructured form to helical structures of \(\sim 30\)–40\% helicity, thus suggesting a strong underlying secondary-structure propensity. Questions remain on how the fold propensity influences the binding mechanism and whether the process is one of conformational selection, induced fit, or a hybrid scenario. A leverage of gaining a better understanding is therapeutic discovery of peptide-based inhibitors.

A recent simulation study of the conformational heterogeneity of the VP35 peptide was reported by applying CHARMM-based potential energy functions and a generalized Born (GB) solvent model. The study found that the extent of conformational plasticity along the secondary-structure forming reaction coordinate depended on the force field and sampling algorithm. In the best outcome, the conformational ensemble of the peptide was determined to be a large manifold of thermally accessible states, which is characteristic of the observed CD disordered ensemble. The less favorable outcome was kinetic trapping of helical conformational states. Regardless of the selected force field, all simulation models tested found the highly populated states to exhibit collapsed measures of radius of gyration (\(R_g\)).

This brief work extends the analysis of the VP35 peptide and reports a simulation study based on a reductionist strategy of modeling the peptide by a coarse-grained (CG) model. The advantage of CG models is a significant reduction in...
computational time in generating structural ensembles as compared to fully atomistic explicit solvent simulations. Current strategies of CG models are quite broad and vary in their atomic granularity of modeling peptides and proteins.\textsuperscript{8} The most popular approach is the molecular dynamics simulation method of applying CG force fields. An alternate to continuous representations and a computationally much faster method is lattice-based polymer physics models. Formulations range from representations and a computationally much faster method is

An intermediate resolution lattice model is the side-chain-only (SICHO) model constructed of a single pseudoatom per residue.\textsuperscript{6,11,12} As pointed out by Kolinski and co-workers in a recent review of CG models and their applications,\textsuperscript{6} the SICHO model has not been extensively studied, despite promising preliminary results.\textsuperscript{13–15} Here, we apply the SICHO model to the VP35 peptide using replica-exchange lattice Monte Carlo simulations. Accurate reconstruction of all-atom models from lattice generated conformations is carried out by a technique developed by Feig and co-workers.\textsuperscript{16} Previous application of the SICHO model with parallel tempering and the rebuilding of all-atom structures is illustrated by the multiscale refinement of protein loops.\textsuperscript{17} Here, the issue is whether the terms of the SICHO potential energy function can be weighted to avoid kinetic traps introduced by a starting helical topology and at the same time allow large-scale conformational heterogeneity of unfolded and folded states. Heterogeneity should include the underlying propensity in the secondary structure as observed by CD with the addition of TFE.

As with previous parallel-tempering simulations of the VP35 peptide, conformational sampling of transitions is assessed by potentials of mean force (PMFs) as a function of fractional helicity ($f_{16}$) and values of $R_g$. What is new here from earlier studies is an application of a CG elastic network interpolation model (ENI) developed by Kim and co-workers\textsuperscript{8–21} to estimate the transition-state cost of excursions on a conformational network. This analysis combined with the PMF suggests a possible hybrid binding process of conformational selection and an induced fit for molecular association with the Ebola virus nucleoprotein. Finally, a comparison is presented of sampling distributions extracted from previous all-atom simulations using a statistical potential for assessing peptide folds.

2. RESULTS AND DISCUSSION

2.1. Conformational Landscape. The advantage of lattice protein models in comparison with all-atom representations is the reduced computational demand to obtain exhaustive conformational sampling and provide efficient characterization of the phase diagram. While sampling is more robust, the issue is the accuracy of CG potential energy functions to correctly generate known conformations and populate the energy landscape. An example of a previous application of the SICHO model is the modeling of protein loops, where the pseudoatoms of the protein stem outside of the loop region were tethered to their initial lattice positions.\textsuperscript{17} The prior study found that, depending on the loop environment, the SICHO model sampled native-like conformational states among the ensembles and their detection depended on the density probability distributions.

Unlike loop modeling, the computational exploration of IDPs is represented as unrestrained sampling and thought to take place on flat energy surfaces. Consequently, the difficulty arises from the inherent biases in potential energy functions that favor the formation of secondary-structure elements. An illustrative example of the effect of biases on secondary structure formation is given by the modeling of the VP35 peptide using different CHARMM-based force fields.\textsuperscript{6,7} The kinetic trapping of states in helical topologies was found to be influenced by the force field and the conformational sampling algorithm. Using a parallel tempering method, the force field CHARMM22+CMAP (designated as C22) distinguished itself from CHARMM36m (C36m)\textsuperscript{12} in producing greater conformational plasticity along the helix-forming reaction coordinate, although the refined C36m produced greater diversity of helical, $\beta$-hairpins, and coils when applied with an adaptive tempering method.\textsuperscript{7}

To avoid the pitfalls of falling into a kinetic trap on the unfolding–folding landscape, the SICHO potential term that favors the formation of regular protein secondary structure elements was scaled in attempt to replicate a set of minimum energy conformers taken from the C22 simulations.\textsuperscript{7} It is found that the lattice simulation with the stiffness parameter $\alpha_{\text{stiffness}}$ set to a value of 1.0, $\alpha_{\text{short}} = 0.5$, and $\alpha_{\text{entropy}} = 0.1$ yielded the best agreement with the C22 energy landscape. Alternative values explored for $\alpha_{\text{stiffness}}$ were set as 0.5 and 0.0, where the latter produced no helical structures.

Shown in Figure 1A,B are the PMFs at a reduced temperature $T = 1$ calculated by the parallel tempering weighted analysis method (PTWHAM) algorithm.\textsuperscript{23} The probability density profiles are denoted as $W_f(f_{16}R_g)$ and $W_f(f_{16}Z)$ for order parameters $f_{16}$, $R_g$, and the SICHO energy $Z$-score. Reconstructed all-atom structures were applied in evaluating $f_{16}$ and $R_g$ from lattice generated conformations. The calculated $W_f(f_{16}R_g)$ reveals the SICHO model produced significant conformational diversity among multiple states, and while the landscape is of lower resolution in defining density contours as compared to all-atom simulations,\textsuperscript{6,7} the SICHO model exhibits conformational specificity in lattice energies as detected in the $W_f(f_{16}Z)$ profile. At the same time, the model avoids kinetic traps to produce PMFs that are consistent with the notion of a heterogeneous ensemble.

As noted in section 4, transitions among helical conformations are with the secondary structure of the bound state annotated in the starting sequence file and the weights of the scaling parameters selected for allowing excursions. The bound peptide shows an initial folded topology of $f_{16} = 0.43$ composed of 4 $10\alpha$-helical conformation for residues Trp28-Gly36 and a 3 $10\alpha$-helix for Val41-Asp43. For problems of modeling peptides without a priori knowledge of the secondary structure, predictions must be made. For this particular peptide, predictions without bias of the crystallographic conformation return an estimate of $f_{16} \approx 0.3$ with probabilities of >0.9 for helical formation in the peptide region of Gly27 to Met34.\textsuperscript{24} Starting with no secondary structure annotated in the sequence leads to rare formation of any helical structures independent of the weights placed on the potential energy function.
A simple measure of sampled space is the analysis of the population density distributions and their free-energy values. The global minimum in $W(f_{H}, R_g)$ is observed at $(f_{H} = 0, R_g = 8.2 \text{ Å})$, and the transition to the weakly populated bound conformation modeled by an approximate PTHWAM bin $(f_{H} = 0.42, R_g = 10.4 \text{ Å})$ yields a $\Delta W_T$ of $4.6k_B T$ with scaling of the reduced $T$ to 300 $K$. To put this magnitude into perspective, the conformational basin that contains the minimum in the SICHO lattice energy ($E_{\text{min}}$) is observed at $(f_{H} = 0.43, R_g = 8.2 \text{ Å})$, and its transition to the bound conformation shows a $\Delta W_T$ of $1.6k_B T$. A similar transition from the $E_{\text{min}}$ to an unfolded conformation shows a $\Delta W_T$ of $1.8k_B T$. A final comparison is the free-energy difference between the minimally folded state given by $(f_{H} = 0.18, R_g = 8.2 \text{ Å})$ and the unfolded state, which yields $\Delta W_T \approx 1k_B T$.

The free energy of shifting a highly populated folded conformation to the global minimum in $W(f_{H}, R_g)$ is observed to be greater than thermal fluctuations and suggests a multistate transition pathway. While Markov state models provide powerful tools to investigate conformational transition networks, their applications to parallel tempering of swapping configurations between ensembles are challenging and require further benchmarking. Alternatively, the ENI model is straightforward, and Figure 1C illustrates its use where the transition-state costs are given by $C_{\text{ENI}}$ for morphing state $i \rightarrow j$. The network was constructed by applying low-energy conformations extracted from clustering and selected modeled structures that minimize $C_{\text{ENI}}$ between states. Results reveal a $C_{\text{ENI}}$ network that is relatively flat across the $f_{H}$ range up to $\approx 0.5$ and displays values that are comparatively lower than the peak value of $C_{\text{ENI}} \approx 4200$ for the transition from a helical-like bundle to the minimum in $W_T(f_{H}, R_g)$. The flatness in transition-state costs combined with free-energy differences near thermal fluctuations is characteristic of a glassy-like protein state.

The unambiguous determination of the mechanism of coupled folding and binding of intrinsically disorder peptides and proteins is a difficult task from either experimental measurements or simulations. Binding mechanisms encompass two different limiting scenarios: (i) conformational selection where a shift in population takes place in an ensemble of unbound states to a bound-like conformation prior to binding and (ii) an induced-fit model where conformational changes accompany the binding processes. From a modeling perspective, the first step in the analysis is to understand the unbound form and the parameters that govern the state populations and their transitions on the free-energy landscape. Here, the implication of the observed conformational plasticity of the peptide up to an $f_{H}$ threshold of 0.5, which include bound-like structures, is a binding mechanism by a process of conformational selection. An observable determined from the simulation is low free-energy barriers for this subset of helical transitions. In contrast, the fold topology in the C-terminal region of the peptide chain that contains the $3_{10}$-helix is weakly populated within the SICHO potential energy function and disappears in sampling with $\sigma_{\text{affinity}} = 0.5$. The $\Delta W_T$ for the transition from the SICHO $E_{\text{min}}$ to the helical-like bundle $(f_{H} = 0.75, R_g = 8.2 \text{ Å})$ is unfavorable by $2.4k_B T$ and exhibits a $C_{\text{ENI}}$ of nearly twice the transition $E_{\text{min}} \rightarrow$ bound state. Because of the high transition barrier, the inference is kinetics that favors a process of induced fit of the helical fold in the C-terminal region. To obtain a more complete picture of binding, the interaction details of the fit requires molecular docking to the viral nucleoprotein and simulations to observe conformational reorganization upon binding. Given the caveat of only modeling the unbound state, the net mechanism of binding is proposed to be a hybrid process.

Figure 2 displays the root-mean-square deviation (RMSD) of the main-chain $C\alpha$ of all-atom conformations reconstructed from the lattice simulation at the replica client $T = 1$. The conformational deviation is from the starting bound fold of the peptide. The plot demonstrates that the lattice Monte Carlo simulation achieved convergence in generating a statistical average $C\alpha$-RMSD of $6.4 \text{ Å}$ with a deviation of $\pm 0.8 \text{ Å}$. In addition, the lattice model reports a sampled RMSD space beyond a narrow range, showing a significant spread and revisiting in conformational space to find a minimum RMSD of 5693 https://dx.doi.org/10.1021/acsomega.9b03581
ACS Omega 2020, 5, 5691–5697
Figure 2. The plot displays the root-mean-square deviation (RMSD) of the main-chain Ca of conformations extracted from the lattice simulation at replica client \( T = 1 \). The number of extracted conformations consisted of 200,000 peptide structures. The RMSD is from the starting bound conformation of the peptide and reports a statistical average for Ctr-RMSD of 6.4 ± 0.8 Å. Illustrated are diverse peptide topological folds that represent the statistical average from a simple moving average.

\(~4\) Å and an extreme of \(~10\) Å. As illustrated in Figure 2, the average RMSD contains a coexistence of a diverse set of peptide folds, ranging from a small helix to an extended helix in the N-terminal region.

2.2. Comparison of the CG Model and CHARMM-Based Distributions. Figure 3A illustrates a contrast of the SICHO model simulation results with PMF data extracted from previously reported all-atom simulations of the peptide using the C36m and C22 force fields with a GB solvent model.6,7 Two calculations are presented for the CG model: (i) results of extracting values from \( W_f(f_{1H}R_g) \) along a \( f_{1H} \) pathway of connecting PTWHAM bins that contain minimum free-energy values and (ii) a pathway of connecting the \( f_{1H} \) bins along a constant \( R_g \) pathway where the value of \( R_g \) is taken from the bin \( \langle f_{1H}R_g \rangle = 0 \). For C36m and C22, the calculations follow the second approach as with the CG model and provide comparative boundaries. Uncertainty in the free energies is \(~0.2k_B T\), determined from different bin sizes in constructing weighted histograms and the application of the multistate Bennett acceptance ratio method25 to test numerical convergence. Errors are greater where populations are limited due to sampling fold topologies that are weakly favorable in the SICHO function.

The plot of Figure 3A shows that, even though on a fixed \( R_g \) path, C22 displays broad helical plasticity among low-free-energy states, and \( W_f(f_{1H}R_g = 8.8\) Å) \( \approx 0 \) takes place at \( f_{1H} = 0.36 \) (recall the bound form exhibits \( f_{1H} = 0.43 \). In stark contrast, C36m reveals the kinetic trapping of a helical fold at \( f_{1H} = 0.26 \) following a downhill slope from the unfolded peptide state and proceeded by a sharp incline to form additional helicity at a significant free-energy cost. Between these two bounds, the CG model with an unrestrained \( R_g \) path best mirrors C22 with the display of coexistence between states of varying fractional helicity at free-energy values near or lower than thermal fluctuations. Despite the coarse-grained treatment, the rebuilding of lattice conformations to all-atom representations obtained sufficient resolution to capture disorder—order transitions among the helical folds.

Illustrated in Figure 3B are plots of helix formation as a function of \( R_g \) evaluated across the 24 replica clients. Shown are statistical averages over datasets for the CG model and simulations applying C22 and C36m taken from examination of earlier data.6,7 The analysis shows that the CG model exhibits the weakest cooperativity among the simulation models. What is furthermore important are the initial slopes of the \( f_{1H} \) profiles in terms of modeling helix stiffness, where both the CG and C22 models show sharp declines, while C36m presents a slow initial drop in \( f_{1H} \). This qualitative ranking of helix stiffness is consistent with the observed plasticity along the helix-forming reaction coordinate shown in Figure 3A. While the statistical average of \(~30\%) helicity produced by C36m would appear to be a result of the refined force field aimed at reducing secondary-structure biases and would seem to be more consistent with CD experiments with TFE, C36m failed to capture significant disorder—order helical fluctuations. Conversely, the CG model of SICHO sample significant helical transitions, and although the statistical average is greater than CD experiments, weighting of the populations across multiple low-free-energy states in \( W_f(f_{1H}R_g) \) is more consistent with the notion of an IDP.

A further observation of the \( f_{1H} \)–\( R_g \) profile for the CG model is that the sampled conformations are more collapsed, approaching the unfolded state than the models C22 and C36m. This result is the outcome of the parametrization of the terms \( \alpha_{\text{stiffness}} \) and \( \alpha_{\text{short}} \) along with \( \alpha_{\text{entropy}} \) to allow the CG model to generate an ensemble of conformational transitions. While the \( R_g \) is lower for the CG model at replica clients containing higher temperatures, values for these extended folds correctly encompass a negligible secondary structure and provide a sufficiently broad statistical distribution for modeling the PMF in Figure 1A.

It is also observed that all simulation models at the lowest client in the exchange show native states that are overly collapsed in comparison to a typical unfolded 28-residue peptide showing an \( R_g \approx 13 \) Å.26 A similar effect is depicted of applying an explicit/implicit solvent hybrid replica-exchange simulation.
method\textsuperscript{27} of modeling the peptide to determine the effect of explicit water interactions.\textsuperscript{6} These combined studies note the inherent deficiencies of all-atom force fields and solvent models that tend to overstabilize fold propensities. Conceivably, alternative CG models CABS\textsuperscript{10}, MARTINI\textsuperscript{28}, and UNRES\textsuperscript{29} may avoid some of these problems. For all-atom simulations, recent developments in molecular mechanics force fields (e.g., refs 30–32) offer wide-ranging parameterizations to reproduce more accurately fold propensities and should help improve modeling the temperature dependence of IDPs.\textsuperscript{33}

Figure 4 shows a distinction of the CG models with datasets extracted from C36m and C22 simulations. The two CG models are with the stiffness parameters ($\alpha_{\text{stiffness}}$) set to 1.0 and 0.5. Profiles are constructed from 5000 conformations and are evaluated using the statistical potential DFIRE\textsuperscript{34} to assess measures of the minimum end-to-end distance of the peptide (denoted as $d_{min}$) and the fraction of native contacts, Q. The X-ray crystallographic conformation shows a DFIRE energy value of -36 kcal/mol and a $d_{min}$ of 22.5 Å.

A comparison of the two CG simulations (Figure 4A,B) shows as anticipated that the lower stiffness value yields a profile of $d_{min}$ that populates conformations with a more opened fold topology. While the sampling range of $d_{min}$ values is satisfactory, both simulations show limited Q values of populating the bound topology. As noted above, the principal reason of selecting $\alpha_{\text{stiffness}} = 1$ is the better sampling of the short helix in the C-terminal half of the peptide. Examples are shown at the minimum energy in DFIRE and the maximum values of Q.

Figure 4C,D shows profiles for C36m and C22. Unlike the CG model, both CHARMM-based simulations only weakly populated values of $d_{min}$ near the bound state. A further distinction is that the all-atom simulations appear to “refine” conformations in terms of DFIRE values that are much more favorable. Among the simulations, C22 samples greater Q values, and the energy minimum exhibits a conformation with a helical topology in the C-terminal region, a fold most similar to the bound state from X-ray diffraction.

3. CONCLUSIONS

This work presented an application of a coarse-grained model in replica-exchange Monte Carlo simulations of a 28-residue peptide extracted from the Ebola virus VP35 protein. With a selection of the weights in the SICHO potential function, the simulation of the peptide produced a free-energy landscape that displayed conformational diversity along the helix-forming reaction coordinate. Despite the diversity in populations, the generated conformational ensemble produced states that showed compacted folds unlike the bound conformation. To better understand the ensemble, an elastic network interpolation model was applied to culled peptide structures and estimate the transition costs of morphing state $i \rightarrow j$. The analysis showed a relatively flat manifold of transition costs of helix formation along the N-terminal end of the peptide and weak helical populations in the C-terminal end. While determination of the binding mechanism is a challenge from either experiments or simulations, the SICHO model suggests from analysis of the unbound transition network a hybrid mechanism through conformational selection of the 4\textsuperscript{10}-helix followed by an induced fit of the 3\textsuperscript{10}-helix. Finally, a comparison of the model simulations with previously reported CHARMM-based simulations of the peptide demonstrates distinctions in the PMFs along helix-forming paths. By applying a statistical potential, the all-atom simulations appear to refine conformations to lower-energy distributions, as anticipated from refined force fields. Among the model simulations, CHARMM22 results are distinct in populating a higher fraction of native contacts, although the SICHO model performed admirably for an intermediate resolution representation. A promising application of the CG model is generating starting conformations for refinement by either all-atom simulations or a combined hybrid approach.

4. COMPUTATIONAL METHODS

4.1. Lattice Ensemble Tempering.

Chain conformations were generated on a cubic lattice using the MONSSTTR program developed by Skolnick and co-workers.\textsuperscript{11,12} In the SICHO model, each amino acid is represented by a single virtual particle located at the side-chain center of mass and projected onto the cubic lattice. The SICHO potential energy function

\[
E_{\text{SICHO}} = \frac{1}{2} \sum_{i,j} K \cdot (r_{ij} - d_{ij})^2
\]

where $r_{ij}$ is the distance between particles $i$ and $j$, and $d_{ij}$ is the distance for the native state. The stiffness parameters $K$ are chosen to achieve the desired level of conformational diversity.

\[
E_{\text{DFIRE}} = \sum_{i,j} \frac{1}{2} \left( \frac{d_{ij} - d_{eq}}{d_{eq}} \right)^2
\]

where $d_{eq}$ is the equilibrium distance, and $d_{ij}$ is the current distance. The DFIRE potential is used to assess the quality of the generated conformations.

\[
E_{\text{bind}} = \sum_{i,j} \frac{1}{2} \left( \frac{d_{ij} - d_{min}}{d_{min}} \right)^2
\]

where $d_{min}$ is the minimum end-to-end distance. The binding energy is used to assess the ability of the model to populate bound states.

\[
E_{\text{score}} = E_{\text{SICHO}} + E_{\text{DFIRE}} + E_{\text{bind}}
\]

The score function is used to rank the generated conformations and select the best ones for further analysis.
incorporates a combination of statistical and empirical terms and takes on the following form

\[ V_{\text{SICHO}} = \alpha_{\text{stiffness}} \sum_{i} V_{\text{stiffness},i}^{\text{stiffness}} + \alpha_{\text{short}} \sum_{i} V_{\text{short},i}^{\text{short}} + \alpha_{\text{pair}} \sum_{i,j} V_{\text{pair},i,j}^{\text{pair}} + \alpha_{\text{3-body}} \sum_{i,j,k} V_{\text{3-body},i,j,k}^{\text{3-body}} + \alpha_{\text{multibody}} \sum_{i,j,k} V_{\text{multibody},i,j,k}^{\text{multibody}} + \alpha_{\text{bilateral}} \sum_{i} V_{\text{bilateral},i}^{\text{bilateral}} + \alpha_{\text{contact}} \sum_{i} V_{\text{contact},i}^{\text{contact}} + \alpha_{\text{centrosym}} V_{\text{centrosym}} \]  

(1)

where the additive $V$ terms are atom-specific potential functions, and the $\alpha$ parameters denote scaling factors. The stiffness parameter $\alpha_{\text{stiffness}}$ controls the scaling of the potential term favoring the formation of secondary-structure elements of the peptide chain. The parameter $\alpha_{\text{short}}$ scales short-range interactions that depend on the type of secondary-structure elements in the sequence file calculated from an algorithm such as DSSP.\(^{35}\) The remaining potentials account for $N$-body interactions of pairwise, long-range, and soft-core repulsive interactions. The scaling parameter $\alpha_{\text{centrosym}}$ sets the scaling of the centrosymmetric potential that is used to favor more compact structures over extended forms.

The peptide sequence is 20-MPGPELSGWISEQMLT-GRIPvSDIFCDI-47, and the folded form of the peptide was extracted from PDB 4YPI.\(^{4}\) The grid size for the cubic lattice was set at a value of 125 lattice units in each direction at a resolution of 1.45 Å grid spacing. Three separate simulations were conducted, where the parameter $\alpha_{\text{stiffness}}$ was varied. One simulation had the parameter set to a value of 1.0 (where the default is 1.25), another set to 0.5, and the last model set to 0.0. Other potential energy function parameters were set at $\alpha_{\text{short}} = 0.5$ (default: 0.5) and $\alpha_{\text{centrosym}} = 0.1$ (default: 0.9). Each simulation was started from the PDB conformation of the bound peptide fold, and the sequence file was annotated with the DSSP secondary structure.

The number of lattice parallel-tempering simulation cycles at each temperature was set to 20, and the number of Monte Carlo moves per cycle was 50. Culled conformations from the simulations were extracted from a sampled population using 24 replicas. Replicas were exponentially spaced from a reduced temperature $T$ of 1.0 to 2.4, where $T$ is normalized by a reference temperature such that $\beta^{-1} = k_{B}T$ represents the energy unit (where $k_{B}$ is the Boltzmann constant). The temperatures for parallel tempering were selected based on a geometrically spaced set. The value $T = 1$ is set to represent the distribution of conformations modeled by the SICHO force field at approximately 300 K. All-atom structures were reconstructed from the lattice simulations by using the Multiscale Modeling Tools for Structural Biology (MMTSB).\(^{36}\) Each structure was refined by energy minimization of applying the C36m force field with the GBMV2 implicit solvent model using a procedure given in the previous study of loop predictions.\(^{37}\) PMFs were calculated using the PTWHAM algorithm\(^{35}\) with order parameters set at modeling $f_{10}$, $R_{g}$ and the lattice energy Z-score. The extracted conformations were clustered using a $k$-means algorithm documented in MMTSB.

Comparison of the SICHO lattice model with data sets from C22 and C36m simulations presented in a previous study centers on computing additional order parameters. These include distributions of the minimum distance between residues Met20 and Ile47, the fraction of side-chain contacts equivalent to the native fold (denoted as $Q$), and scoring conformations using the statistical potential DFIRE.\(^{34}\) The simulations with C22 and C36m were conducted using the GBMV2 solvent model, and details of the methods are given in the previous studies.\(^{6,7}\) Previous studies also include the application of an adaptive temperature-based replica-exchange simulation method in improve conformational sampling.\(^{37}\)

4.2. Elastic Network Model. Conformational transitions between clustered states of the VP35 peptide are further evaluated by applying the ENI model.\(^{18-21}\) Calculations morph state $i$ to state $j$ and start by transforming the coordinates for two peptide conformations into a non-standardized form by adjusting the center of mass and principal axis for each molecule. Rather than applying internal coordinates, ENI converts two sets of distances between spatially close representative atoms along the two chains. The coordinates for the starting conformer is denoted by $\{x_{i}\}$ and the final conformer $\{x_{j}\}$ along a pathway. Intermediate conformations are generated iteratively by solving a quadratic potential-like cost function given by

\[ C(\delta) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i+1}^{N} k_{ij} (\|x_{i} + \delta - x_{j} - \delta\| - I_{ij})^{2} \]  

(2)

where $k_{ij}$ is a spring constant, $N$ is the number of representative atoms, and $\delta$ is a 3N-dimensional vector of displacements. The term $I_{ij}$ is the desired length at a certain intermediate state determined by distance interpolation and is given by

\[ I_{ij} = (1 - \alpha) \|x_{i} - x_{j}\| + \alpha \|x_{i} - x_{j}\| \]  

(3)

where $\alpha$ is the coefficient specifying how far a given state is along the transition from the starting conformer to the final conformer. When the coefficient $\alpha$ is 0.5, the calculated conformation is positioned with inter-residue distances at the average of conformations $\{x_{i}\}$ and $\{x_{j}\}$. For the interpolation scheme applied here, analyzed conformations were extracted from the lattice replica exchange at $T = 1$ in the PDB format as an input to KOSMOS.\(^{14}\) The query option in KOSMOS was set to advance level with coarse-graining performed at the Cα scale, and the cutoff distance was set at 50 Å.

## AUTHOR INFORMATION

### Corresponding Author

**Mark A. Olson** – Systems and Structural Biology Division, USAMRIID, Frederick, Maryland 21702, United States; orcid.org/0000-0002-5259-0343; Email: mark.a.olson1.civ@mail.mil

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.9b03581

### Notes

The author declares no competing financial interest.

## ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Defense Threat Reduction Agency grant (4.100011_07_RD_B). The opinions or assertions contained herein belong to the author and are not necessarily the official views of the U.S. Army or the U.S. Department of Defense.
REFERENCES

(1) Leung, D. W.; Borek, D. D.; Luthra, P.; Binning, J. M.; Anantpadma, M.; Liu, G.; Harvey, I. B.; Su, Z.; Endlich-Frazier, A.; Pan, J.; Shahman, R. S.; Chiu, W.; Davie, R. A.; Otwinowski, Z.; Basler, C. F.; Amarasinghe, G. K. An intrinsically disordered peptide from Ebola virus VP35 controls viral RNA synthesis by modulating nucleoprotein-RNA interactions. Cell Rep. 2015, 11, 376–389.

(2) Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T. Finding our way in the dark proteome. J. Am. Chem. Soc. 2016, 138, 9730–9742.

(3) Perdiguó, N.; Heinrich, J.; Stolte, C.; Sabir, K. S.; Buckley, M. J.; Tabor, B.; Signal, B.; Gloss, B. S.; Hammang, C. J.; Rost, B.; Schafferans, A.; O’Donoghue, S. I. Unexpected features of the dark proteome. Proc. Natl. Acad. Sci. U. S. A. 2015, 112, 15898–15903.

(4) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 2015, 16, 18–29.

(5) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing protein intrinsic disorder. Chem. Rev. 2014, 114, 6561–6588.

(6) Olson, M. A. On the helix propensity in generalized Born solvent descriptions of modeling the dark proteome. Front. Mol. Biosci. 2017, 4, 3.

(7) Olson, M. A. Parallel tempering of dark matter from the Ebola virus proteome: Comparison of CHARMM36m and CHARMM22 force fields with implicit solvent. J. Chem. Inf. Model. 2018, 58, 111.

(8) Kniecik, 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.

(9) Shakhnovich, E. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chem. Rev. 2006, 106, 1559–1588.

(10) Kolinski, A. Protein modeling and structure prediction with a reduced representation. Acta Biochim. Pol. 2004, 51, 349–371.

(11) Skolnick, J.; Kolinski, A.; Ortiz, A. MONSTER: A method for folding globular proteins with a small number of distance restraints. J. Mol. Biol. 1997, 265, 217–241.

(12) Kolinski, A.; Skolnick, J. Assembly of protein structure from sparse experimental data: An efficient Monte Carlo model. Proteins 1998, 32, 475–494.

(13) Boniecki, M.; Rotkiewicz, P.; Skolnick, J.; Kolinski, A. Protein fragment reconstruction using various modeling techniques. J. Comput. Aided Mol. Des. 2003, 17, 725–738.

(14) Stumpf-Kane, A. W.; Maksimiak, K.; Lee, M. S.; Feig, M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. Proteins: Struct., Funct., Bioinf. 2008, 70, 1345–1356.

(15) Kihara, D.; Zhang, Y.; Lu, H.; Kolinski, A.; Skolnick, J. Ab initio protein structure prediction on a genomic scale: application to the Mycoplasma genitalium genome. Proc. Natl. Acad. Sci. U. S. A. 2002, 99, 5993–5998.

(16) Feig, M.; Rotkiewicz, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L., III Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models. Proteins 2000, 41, 86–97.

(17) Olson, M. A.; Feig, M.; Brooks, C. L., III Prediction of protein loop conformations using multiscale modeling methods with physical energy scoring functions. J. Comput. Chem. 2008, 29, 820–831.

(18) Lee, H.; Seo, S.; Kim, M.; Choi, J. B.; Kim, S. M.; Jeon, T.-J.; Kim, M. K. Opening and closing of a toroidal group II chaperonin revealed by a symmetry constrained elastic network model. Protein Sci. 2014, 23, 703–713.

(19) Seo, S.; Kim, M. K. KOSMOS: a universal morph server for nucleic acids, proteins and their complexes. Nucleic Acids Res. 2012, 40, W531.

(20) Kim, M. K.; Jernigan, R. L.; Chirkjian, G. S. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. Biophys. J. 2005, 89, 43–55.

(21) Kim, M. H.; Kim, Y. J.; Kim, H. R.; Jeon, T. J.; Choi, J. B.; Chung, K. Y.; Kim, M. K. Computational simulation of the activation cycle of gp subunit in the gp protein cycle using an elastic network model. PLoS One 2016, 11, e0159528.

(22) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat. Methods 2017, 14, 71–73.

(23) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. J. Chem. Theory Comput. 2007, 3, 26–41.

(24) Kieslich, C. A.; Smadbeck, J.; Khoury, G. A.; Floudas, C. A. conSSert: Consensus SVM model for accurate prediction of ordered secondary structure. J. Chem. Inf. Model. 2016, 56, 455–461.

(25) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 2008, 129, 124105–124110.

(26) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; et al. Random-coil behavior and the dimensions of chemically unfolded Proteins. Proc. Natl. Acad. Sci. U. S. A. 2004, 101, 12491–12496.

(27) Chaudhury, S.; Olson, M. A.; Tawa, G.; Wallqvist, A.; Lee, M. S. Efficient conformational sampling in explicit solvent using a hybrid replica exchange molecular dynamics method. J. Chem. Theory Comput. 2012, 8, 677–687.

(28) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. J. Chem. Theory Comput. 2013, 42, 6801–6822.

(29) Liwo, A.; Czaplewski, C.; Oldziej, S.; Rojas, A. V.; Kazmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Schera, H. A. Simulation of protein structure and dynamics with the coarse-grained UNRES force field. In Coarse-Graining of Condensed Phase and Biomolecular Systems; Voth, G. G.; Taylor & Francis, 2008, Chapter 8, pp 107–122.

(30) Liu, H.; Song, D.; Lu, H.; Luo, R.; Chen, H.-F. Intronically disordered protein-specific force field CHARMM36IDPFF. Chem. Biol. Drug Des. 2018, 92, 1722–1735.

(31) Song, D.; Wang, W.; Ye, W.-j.; Luo, R.; Chen, H.-F. fIDP4 force field improving the conformation sampling of intrinsically disordered proteins. Chem. Biol. Drug Des. 2017, 89, 5–15.

(32) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. Proc. Natl. Acad. Sci. U. S. A. 2011, 108, E4758–E4766.

(33) Jephthah, S.; Staby, S.; Kragelund, B. B.; Skepø, M. Temperature dependence of intrinsically disordered proteins in simulations: What are we missing? J. Chem. Theory Comput. 2019, 15, 2672–2683.

(34) Zhou, H.; Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci. 2002, 11, 2714–2726.

(35) Kabcsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983, 22, 2577–2637.

(36) Feig, M.; Karanicolas, J.; Brooks, C. L., III MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. J. Mol. Graphics Modell. 2004, 22, 377–395.

(37) Lee, M. S.; Olson, M. A. A comparison of two adaptive temperature-based replica exchange methods applied to a sharp phase transition of protein unfolding-folding. J. Chem. Phys. 2011, 134, 124411–124417.