12-31-2018

Domain Rearrangement and Denaturation in Ebola Virus Protein VP40

Rudramani Pokhrel
Florida International University

Pornthep Sompornpisut
Chulalongkorn University

Prem Chapagain
Florida International University

Brian Olson
University of Southern Mississippi

Bernard Gertsman
Florida International University

See next page for additional authors

Authors

Rudramani Pokhrel, Pornthep Sompornpisut, Prem Chapagain, Brian Olson, Bernard Gertsman, and Ras B. Pandey

Domain rearrangement and denaturation in Ebola virus protein VP40

Rudramani Pokhrel,[1] Pornthep Sompornpisut,[2] Prem Chapagain,[1]
Brian Olson,[3] Bernard Gerstman,[1] and R. B. Pandey[4,a]

[1]*Department of Physics, Florida International University, Miami, Florida 33199, USA*
[2]*Department of Chemistry, Chulalongkorn University, Bangkok 10330, Thailand*
[3]*School of Polymer and High Performance Materials, University of Southern Mississippi, Hattiesburg, Mississippi 39406, USA*
[4]*School of Mathematics and Natural Sciences, University of Southern Mississippi, Hattiesburg, Mississippi 39406, USA*

The VP40 protein plays a critical role in coordinating the virion assembly, budding, and replication of the Ebola virus. Efforts have been made in recent years to understand various aspects of VP40 structure, dynamics, and function such as assembly of the protein and its roles in virus replication and penetration of the protein into the plasma membrane. A major conformational transformation is necessary for VP40 to form some of its oligomeric structures and to perform various functions. This conformational change from a compact structure with the N-terminal domain (NTD) and C-terminal domain (CTD) closely associated involves a dissociation or springing-out of the CTD from the NTD. We perform investigations using computational molecular dynamics simulations as well as knowledge-based Monte Carlo simulations. We find that a sharp springing of the CTD from the NTD in a free VP40 protein cannot occur solely by random thermal fluctuations without intermediate oligomerized segments, and therefore is likely triggered by additional molecular events. © *2018 Author(s).*
All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
https://doi.org/10.1063/1.5063474

INTRODUCTION

The VP40 protein plays a critical role in coordinating the virion assembly, budding, and replication of the Ebola virus.[1–6] Enormous efforts have been made in recent years to understand various aspects of VP40 structure, dynamics, and function such as assembly of the protein[1] and its roles in virus replication[2] and penetration of the protein into the plasma membrane.[3] These molecular investigations on VP40 may help in designing therapeutics[4] against the Ebola virus. As part of these investigations, it is helpful to understand the structure and dynamics of a VP40 monomer in order to assess how its various oligomeric states perform their functions.

A VP40 monomer (Fig. 1) is composed of a distinct C-terminal domain (CTD) connected by a linker segment to an N-terminal domain (NTD). A major conformational transformation is necessary for VP40 to form some of its oligomeric structures and to perform various functions. This transformer conformational change from a compact structure with the NTD and CTD closely associated (Fig. 1a) involves a dissociation or springing-out of the CTD (Fig. 1b) from the NTD. Figure 1a is the initial configuration used in the molecular dynamics computational investigations discussed below. The information for this initial, compact configuration of Fig. 1a was obtained from the Protein Data Bank (PDB ID 1ES6). The crystal structure for the sprung configuration of Fig. 1b is not available.

---

[a]Corresponding author: R.B. Pandey, School of Mathematics and Natural Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA, Telephone: 601 266 4485, Email: ras.pandey@usm.edu
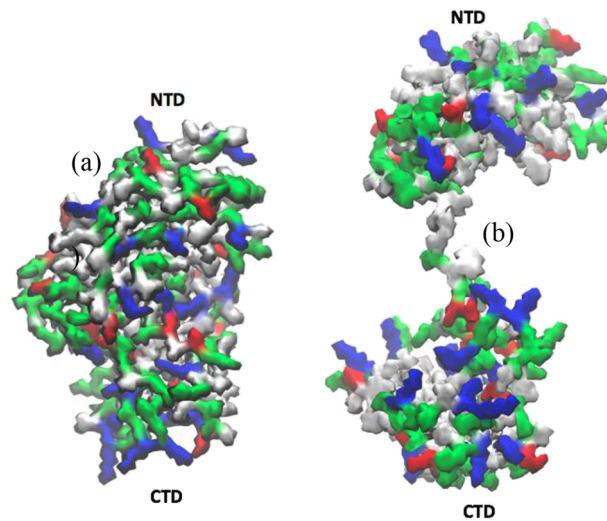
FIG. 1. The VP40 monomer is composed of two domains: NTD and CTD. (a) Protein configuration with the two domains closely associated. (b) Protein configuration with the domains dissociatively "sprung" but still connected by a linker segment.

In order to create Fig. 1b, we pulled gently on the atoms of the CTD to 'spring' the CTD away from the NTD at 300K without disrupting the CTD's folded domain structure. We note that the image in Fig. 1b is used only to pictorially display the 'sprung' configuration and that this arrangement is not used in any of the calculations.

In the sprung configuration, the CTD and NTD remain connected by the linker segment, but have little interfacial contact. The causation of this domain dissociation is under investigation. Silva et al.[1] have pointed out that "VP40 proceeds through intermediate states of assembly (e.g. octamers) but it remains unclear how these intermediates are coordinated with the various stages of the life cycle" particularly the extreme N and C terminals in oligomerization processes. They investigated the 'induced structural and conformational changes' of VP40 particularly aggregation and urea induced denaturation using hydrogen-deuterium exchange mass spectroscopy. They observed only a single conformational transition with an onset concentration of 2 M urea and concluded that "the full length monomer is remarkably stable in presence of high concentration of urea and incapable of self-assembly". With computational modeling it may be easier to analyze the functional conformational springing transition in terms of the interplay between amino acid residue-residue interactions versus thermal agitation. Our investigations make use of computational molecular dynamics simulations as well as knowledge-based Monte Carlo simulations. Data presented in this study shows how the thermal agitations control the oligomerization as the protein undergoes structural transformation. We find that the springing of the CTD from the NTD of a free monomer without bridging via segmental oligomerization cannot occur by random thermal fluctuations alone, and therefore is likely triggered by additional and perhaps specific molecular events as proposed by Ref. 7.

METHODS

In order to investigate large scale structural changes in the VP40 monomer, we use computer simulations over a range of temperatures. By simulating the system at higher temperatures, thermally induced structural changes can occur in time-scales short enough to observe in computational simulations. In all simulations, the initial configuration of the VP40 monomer has the NTD and CTD in close contact.

Molecular dynamics simulations

The initial, compact configuration of the VP40 with the NTD and CTD closely associated was obtained from the Protein Data Bank (PDB ID 1ES6). Missing residues were added using the Modeler

software.[8] We performed coarse-grained molecular dynamic (CGMD) simulations using the Martini force field in which approximately four heavy atoms are mapped onto a single interaction site (bead).[9–11] The monomer system was generated using the Martini-maker plugin in the Charmm-gui web server[12–14] with the Martini22 non-polarizable water model with 0.15 mM NaCl so that counterions were available to neutralize the system. The total system of VP40 monomer and water had a total of 5944 beads. All of our CGMD simulations were run using Gromacs-5.1.1(gpu-version).[15,16] A time step of 15 fs was used for the coarse-grained simulations. The Lennard-Jones and coulomb potential cutoff was set at 12 Å. The Verlet neighbor scheme was used with a straight cutoff to keep track of particles within specified distances. Coulomb interactions were treated using a PME. Ring systems and stiff bonds were controlled using the LINCS algorithm. The pressure was maintained at 1 bar using the Berendsen barostat during equilibration runs and the Parrinello–Rahman barostat during production runs. Pressure coupling was isotropic with a compressibility of 4.5×10^{-4} bar^{-1}. The temperature coupling was done using a velocity rescale algorithm. In order to compare our CG computations with the MC results, as well as to allow conformational changes to occur, we did not restrain the protein with an elastic network. The conformational changes that occur without constraining the secondary structure with the elastic network, as well as the use of the non-polarizable water place limitations on the accuracy of the fine-scale representation of the structural dynamics of the protein.

We ran simulations at eight different temperatures: 300, 450, 550, 600, 700, 800, 1000 and 1100 K. Each system was run for 5 ns of minimization, 200 ns of equilibration, and 500 ns of production run. We used the VMD[17] visualization tool for analysis of our CGMD runs. The analysis for the radius-of-gyration (discussed below) was performed on the final 100 ns of the production run.

Monte Carlo simulations

Monte Carlo (MC) computational simulations involve a coarse-grained representation of amino acids by a node on a cubic lattice with several degrees of freedom.[18,19] Even though the atomic scale structure of each residue is ignored, their specificity is captured via unique residue-specific interactions. The protein VP40 is a chain of 326 residues (Figure S1) tethered together by covalent/peptide bonds; the length of the covalent bond between consecutive residues in the MC simulations varies between 2 and $\sqrt{(10)}$ in units of the lattice constant. The bond-fluctuation approach used here is one of the most efficient and effective methods to investigate large-scale complex problems[20] in polymers, soft matter, and proteins.[18,19] The degrees of freedom for nodes to move and bonds to fluctuate, though limited, are well-defined on a cubic lattice in comparison to off-lattice, continuum space generally used in all-atom MD simulations such as the one used here. The degrees of freedom can be easily enhanced by fine-graining,[19] and allows extension of a computationally efficient[20] approach to protein modeling.

A generalized Lennard-Jones potential[18,19] involving knowledge-based residue-residue interactions[21,22] as input is used for the residue-residue interaction with a range (r_c). It is worth pointing out that the knowledge based residue-residue interactions[21,22] involves a huge ensemble of protein structures from the protein data bank (PDB). Of 400 residue-residue contact elements of a 20×20 matrix of 20 amino acids, there are 210 independent pair interactions. Each interaction pair is unique with attractive (negative) and repulsive (positive) matrix elements[21,22] that capture their specificity. Since the structures (i.e. X-ray snapshots) in the PDB involve proteins in real laboratory samples, interactions of each residue with the solvent and other constituents of the underlying sample are implicitly included in the estimates of the residue-residue contact interactions. Even though the generalized LJ potential[18,19] used here is phenomenological, it captures the specificity of residues somewhat effectively. Each residue performs their stochastic movement guided by the Metropolis algorithm at each temperature (T). A series of attempts to move each residue defines one Monte Carlo step (MCS). Simulations are performed for sufficiently long times (t), typically $t = 10^7$ *MCS* time-steps, so that the system reaches equilibrium. Starting with an initial random configuration, simulations are performed on a 350^3 lattice with *100* independent samples, at each temperature to estimate average values of local and global physical quantities. MC parameters are measured in arbitrary scales (i.e. T in the reduced unit of the Boltzmann constant, sample size in lattice constant, etc.). Connection between the arbitrary reduced units and the realistic units (order of magnitude) can be made by comparing

and calibrating the results of the coarse-grained MC approach with the data from MD simulations (see below). Different lattice sizes are also used to make sure that our results (qualitative trends) are independent of lattice size.

RESULTS AND DISCUSSION

Figure 2 displays the results of the MD simulations. The radius-of-gyration (R_g) of the VP40 monomer is given as a function of temperature. As can be seen in the figure, as the temperature increases, the VP40 monomer undergoes a structural transition from a compact configuration to an extended configuration. Upon examination of the trajectory, we found that the high-temperature extended configuration of the VP monomer was not due to dissociative springing of the NTD from the CTD, but instead was due to denaturation of the protein. In order to better understand the effects of residue-residue interactions versus thermal fluctuations, we analyzed the conformational response using knowledge-based MC simulations.

In order to use the knowledge-based MC simulations together with the MD simulations to understand the dynamics of the VP40 monomer, we first checked if both computational methods resulted in similar behavior. Figure 3 shows that the R_g as a function of temperature displays similar behavior in the MC simulations as it does in the MD simulations. The MD values for temperature and R_g provide approximate calibration of the units for the MC simulations.

Snapshots of VP40 conformations from the knowledge-based MC simulations at representative temperatures are presented in Figures 4 and S2 (supplementary material). These snapshots also show that thermal fluctuations do not lead to springing of the CTD from the NTD, but instead result in segmental oligomerization and denaturation. Clustering of self-assembled residues occurs at low temperatures (i.e., *T = 0.019-0.022*) and gradually reduces as the size of protein expands on raising the temperature. Some degree of domain formation persists at the intermediate temperatures but disappears upon raising the temperature further (see Figure S2 at *T = 0.025*).

Trends in the evolution of segmental domain growth and decay can be better clarified by examining the contact maps of Fig. 5 which shows how the coordinated segmental structures disappear as the temperature is increased. For example, at a low temperature (*T = 0.013*), segmental organization, a cooperative phenomenon resulting from interactions among the neighboring residues along the protein backbone, shows a phase-separation with three distinct regions (^{1}M – ^{120}Y, ^{125}F – ^{223}G, ^{225}K – ^{326}K, Figure S1). Increasing the temperature leads to the appearance of larger loop segments and decay of globular density in the intermediate temperature region, *T=0.021,* which may be due to the onset of denaturation and further investigated below. Note that some degree of segmental globular structure still exists in the center region (involving ^{111}T – ^{228}S) of the protein at *T = 0.021*. All organized structure vanishes at high temperature (*T = 0.025*, see Figure S2).
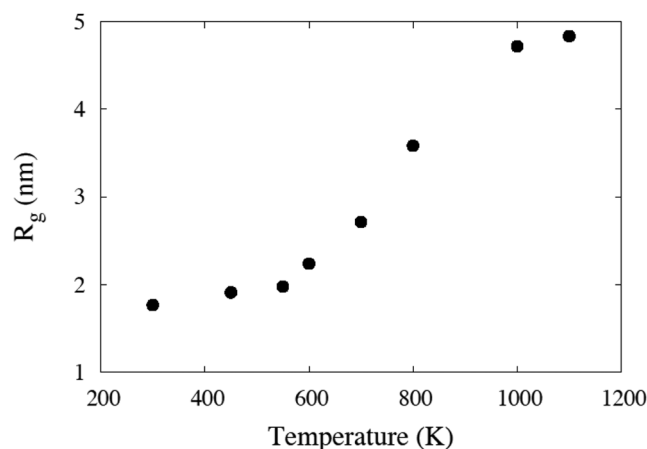

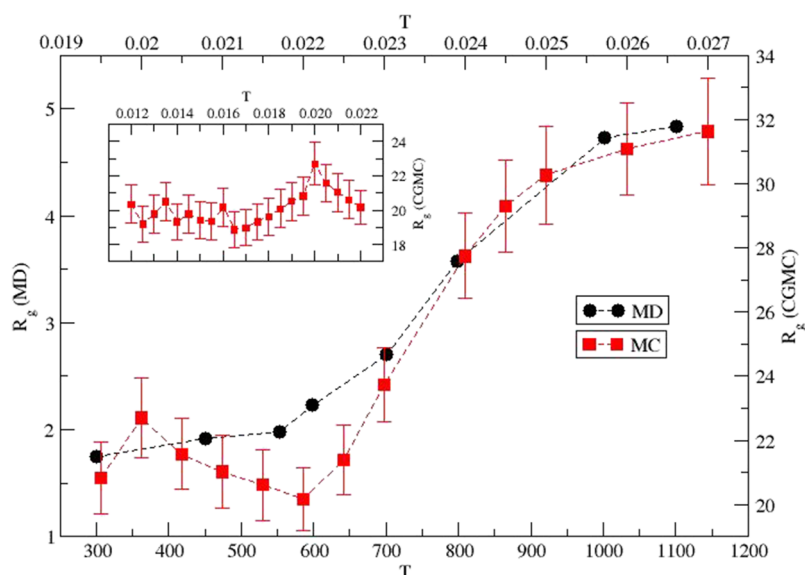FIG. 2. From the MD simulations, the radius-of-gyration of the VP40 monomer as a function of temperature.

FIG. 3. Radius of gyration versus temperature for MD and MC simulations. The X-scale for the temperature is in $^{\circ}$K on the bottom for the MD data and in arbitrary units on the top for MC data. Similarly the Y-axis data from MD simulations (left) are in nm while that from the MC (right) are in arbitrary units. Attempts are made to overlap data points (guide to the eye) from both simulations not only to see if trends in thermal responses are qualitatively similar but also for approximate calibration of the arbitrary units of MC data with that of the MD units. Inset: MC data at especially low temperatures.

Note that the contact map (Fig. 5) and snapshots (Fig. 4 and Fig. S2) represent transient structures. In order to assess the overall contacts, one has to quantify the degree of contacts of each residue and evaluate its average values over many configurations of the protein as follows. The average number N_n of residues around each residue within the range of interaction is a measure of the contact density and provides insight into the segmental stability of the local structures. The contact density profiles are presented in Fig. 6 and Fig. S3 at representative low, medium, and high temperatures. The contact density of the N-tail, residues ^{1}M$-^{13}$Y, the main part of the NTD (^{70}S$-^{185}$P), and the C-tail (^{290}P$-^{326}$K) (Fig. S3) are noteworthy as the extreme N and C terminal tails seem to 'stabilize the monomeric state through a direct association' as reported by Silva et al.[1] and Gc et al.[23] Both Ref. 1 and Ref. 23 discuss how various interactions in the tail segments facilitate domain-association, stabilize

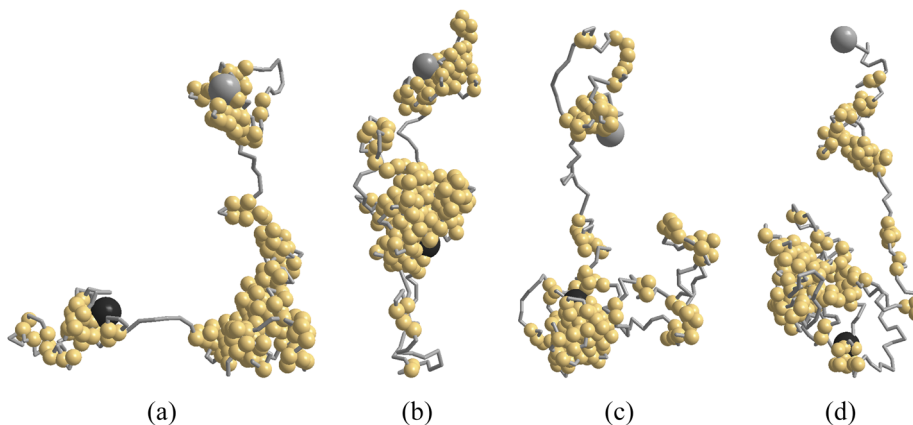FIG. 4. Snapshot of VP40 conformation at the end of 10^7 MC time steps at temperature T = 0.019, 0.020, 0.021, and 0.022 from left to right respectively. Residues close enough to experience interactions are shown by spheres (gold). Large spheres represent end residues (^{1}M (N-terminal in black), ^{326}K (C-terminal in grey)).
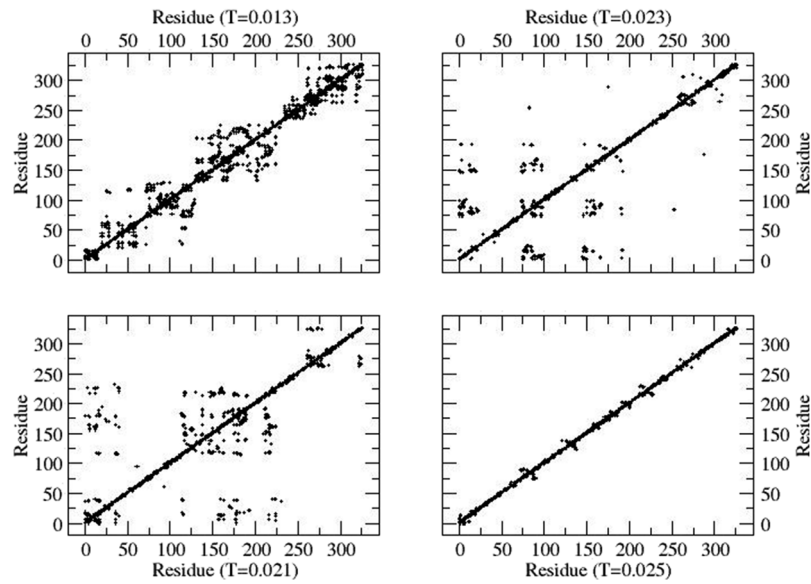
FIG. 5. Contact map of residues in conformations of protein VP40 at representative temperatures.

the domain associated configuration, and play a role in the binding of VP40 to the host cell membrane and the subsequent budding of new virus-like particles. We find that the domains evolve with the temperature as a result of its redistribution and growth and decay of its size (Fig. 6). For example at a low temperature (*T=0.013*, Fig. S3), each of the three clusters (N-tail [1]M–[40]E, NTD [72]A –[195]T, CTD [201]G–[305]M) has a relatively high contact density as the residues self-assemble into segmental globules (see Fig. S2 for the corresponding snapshot). The segmental clusters almost disappear at high temperature (Fig. S2 at *T=0.025*) due to denaturation. Note that the high contact density of the extreme N-tail ([1]M–[13]Y) persists over a relatively large range of temperature (*T=0.013−0.023*) while the globular state of the C-terminal disappears. The large segment of the NTD ([70]S−[185]P) phase-separates into two domains ([72]A−[105]T and [145]H−[185]P) on raising the temperature from *T=0.013* to *T=0.021*. The segmental mobility profile appears to complement the contact map profile, i.e., lower mobility at higher contact density (Fig. S4).
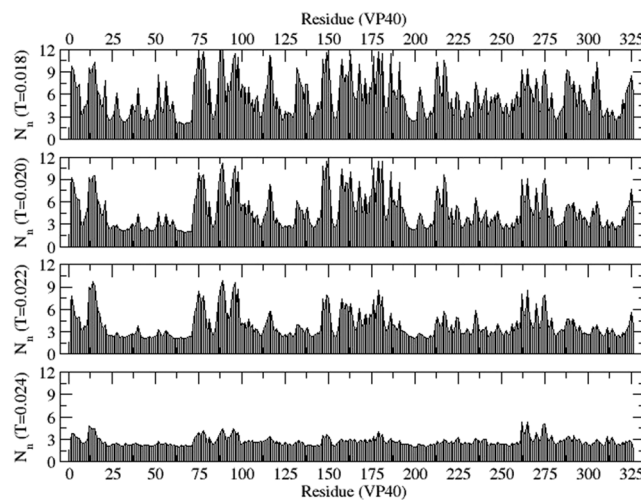

FIG. 6. Average number N_n of residues around each residue within the range of interaction at temperatures T = 0.018, 0.020, 0.022, and 0.024.
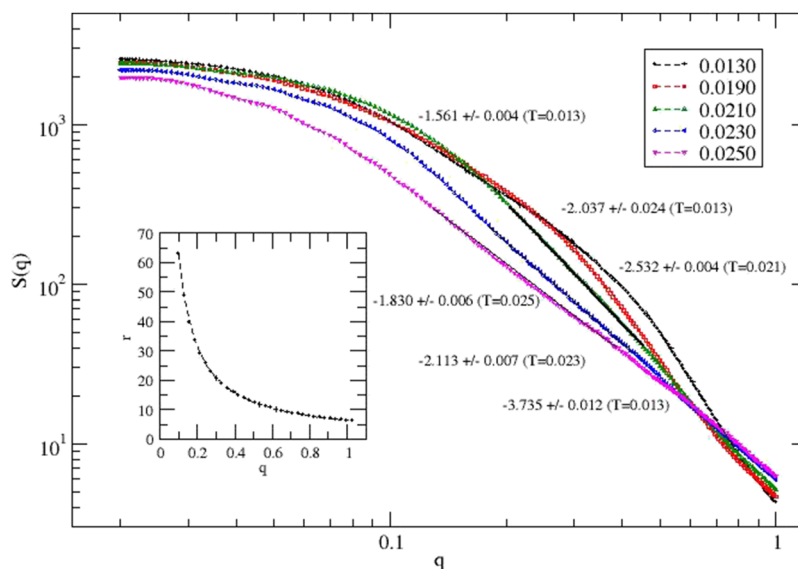
FIG. 7. Variation of the structure factor $S(q)$ with the wave vector q for representative temperatures on a log-log scale. Inset is a guiding scale for the spatial length scale r (for the wavelength which spans over the radius of gyration and the wave vector q). Estimates of the slopes represent $-1/\gamma$.

Denaturation versus domain dissociation can also be quantified through the effective dimension D of the protein chain, which is a measure of how the mass is distributed at various length scales. Compact structures, such as domains have D~3, whereas denatured, random-coil structures have D≤2. The structural dimensionality can be determined[5] by calculating the structure factor $S(q)$ as a function of the wave vector $|q| = 2\pi/\lambda$, with the wavelength λ used for q in Fig. 7 is given in MC lattice units. Variation of the structure factor $S(q)$ with the wave vector q is presented in Fig. 7 for representative temperatures. From Fig. 7, we can determine the scaling of the structure factor $S(q)$ with the wave vector q, $S(q) \propto q^{-1/\gamma}$ to get a value for the exponent. Fitting the data for the structure factor at different length scales (λ) comparable to R_g, one can estimate the effective dimension $D \approx 1/\gamma$ for the mass distribution $N \propto R_g^{1/\gamma}$. Estimates for $-1/\gamma$ at various T and for various length scales are provided in Fig. 7, which shows that the overall protein structure is globular ($D \sim 3$) at the low temperature ($T=0.013$). The effective dimension D decreases on increasing the temperature, i.e., $D \sim 2.5$ at $T= 0.021$, $D\sim 2$ at $T \geq 0.23$ which implies that the conformation of the protein becomes a denatured random- coil at higher temperatures. The globular ($D \sim 3$) to random coil ($D\sim 2$) transition appears to be continuous as a function of temperature.

CONCLUSION

A crucial aspect of the structural-function behavior of the Ebola virus protein VP40 is its ability to dissociate its CTD from its NTD to perform different functions. To determine if this springing apart of the domains occurs due to random thermal motions, we have used both molecular dynamics and knowledge-based Monte Carlo computational simulations to investigate the thermally induced behavior of a monomer of VP40. Conformation of the protein clearly shows that domains appear as the residues self-assemble in globules at low temperatures. Domains are localized around N-tail, the intermediate NTD region, and C-terminal at low temperature, but the domain structures disappear at high temperature where the protein denatured into a random-coil conformation. Therefore, the springing of the CTD from the NTD cannot occur by random thermal fluctuations alone without bridging via segmental oligomerization, and that instead, the protein is more likely to denature. Thus, the functionally important springing apart of the domains is likely triggered by additional mechanism such as a specific molecular event as proposed by Bornholtd.[7]

SUPPLEMENTARY MATERIAL

The supplementary material contains the sequence of protein VP40 which are required to identify relevant amino acids referred in the main body of the manuscript. Additional figures included in the supplement are: snapshots at representative low and high temperatures (Figure S2), average contact measure of residue contact map of residues (Figure S3) for a range of temperatures, and their mobility profile (Figure S4). These are useful in understanding the description provided in the main text of the manuscript.

ACKNOWLEDGMENTS

[1] L. P. Silva, M. Vanzile, S. Bavari, J. M. Aman, and D. C. Schriemer, "Assembly of Ebola virus matrix protein VP40 is regulated by latch-like properties of N and C terminal tails," PLoS One **7**, e39978 (2012).
[2] T. Hoenen, V. Volchkov, L. Kolesnikova, E. Mittler, J. Timmins, M. Ottmann, O. Reynard, S. Becker, and W. Weissenhorn, "VP40 octamers are essential for Ebola virus replication," J Virol **79**, 1898–1905 (2005).
[3] E. Adu-Gyamfi, S. P. Soni, Y. Xue, M. A. Digman, E. Gratton, and R. V. Stahelin, "The Ebola virus matrix protein penetrates into the plasma membrane: A key step in viral protein 40 (VP40) oligomerization and viral egress," J Biol Chem **288**, 5779–5789 (2013).
[4] J. J. Madara, Z. Han, G. Ruthel, B. D. Freedman, and R. N. Harty, "The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target," Future Virol **10**, 537–546 (2015).
[5] J. B. Gc, B. S. Gerstman, R. V. Stahelin, and P. P. Chapagain, "The Ebola virus protein VP40 hexamer enhances the clustering of PI(4,5)P2 lipids in the plasma membrane," Phys Chem Chem Phys **18**, 28409–28417 (2016).
[6] J. B. Gc, B. S. Gerstman, and P. P. Chapagain, "Membrane association and localization dynamics of the Ebola virus matrix protein VP40," Biochim Biophys Acta **1859**, 2012–2020 (2017).
[7] Z. A. Bornholdt, T. Noda, D. M. Abelson, P. Halfmann, M. R. Wood, Y. Kawaoka, and E. O. Saphire, "Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle," Cell **154**, 763–774 (2013).
[8] B. Webb and A. Sali, "Comparative protein structure modeling using MODELLER," Curr Protoc Protein Sci **86**, 2 9 1–2 9 37 (2016).
[9] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, "The MARTINI coarse-grained force field: Extension to proteins," J Chem Theory Comput **4**, 819–834 (2008).
[10] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: Coarse grained model for biomolecular simulations," J Phys Chem B **111**, 7812–7824 (2007).
[11] D. H. de Jong, G. Singh, W. F. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, and S. J. Marrink, "Improved parameters for the Martini coarse-grained protein force field," J Chem Theory Comput **9**, 687–697 (2013).
[12] Y. Qi, H. I. Ingolfsson, X. Cheng, J. Lee, S. J. Marrink, and W. Im, "CHARMM-GUI Martini maker for coarse-grained simulations with the Martini force field," J Chem Theory Comput **11**, 4486–4494 (2015).
[13] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," J Comput Chem **29**, 1859–1865 (2008).
[14] P. C. Hsu, B. M. H. Bruininks, D. Jefferies, P. Cesar Telles de Souza, J. Lee, D. S. Patel, S. J. Marrink, Y. Qi, S. Khalid, and W. Im, "CHARMM-GUI Martini maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides," J Comput Chem **38**, 2354–2363 (2017).
[15] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: Fast, flexible, and free," J Comput Chem **26**, 1701–1718 (2005).
[16] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015).
[17] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," J Mol Graph **14**, 33–38 (1996), 27–38.
[18] S. Kitjaruwankul, C. Khrutto, P. Sompornpisut, B. L. Farmer, and R. B. Pandey, "Asymmetry in structural response of inner and outer transmembrane segments of CorA protein by a coarse-grain model," J Chem Phys **145**, 135101 (2016).
[19] R. B. Pandey and B. L. Farmer, "Aggregation and network formation in self-assembly of protein (H3.1) by a coarse-grained Monte Carlo simulation," J Chem Phys **141**, 175103 (2014).
[20] K. Binder, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford University Press, Place Published, 1995).
[21] S. Miyazawa and R. L. Jernigan, "Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation," Macromolecules **18**, 534–552 (1985).
[22] M. R. Betancourt and D. Thirumalai, "Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes," Protein Sci **8**, 361–369 (1999).
[23] J. B. Gc, K. A. Johnson, M. L. Husby, C. T. Frick, B. S. Gerstman, R. V. Stahelin, and P. P. Chapagain, "Interdomain salt-bridges in the Ebola virus protein VP40 and their role in domain association and plasma membrane localization," Protein Sci **25**, 1648–1658 (2016).