Research article

Thermal-induced unfolding-refolding of a nucleocapsid COVN protein

Warin Rangubpit[1,2], Pornthep Sompornpisut[2], and R.B. Pandey[1,*]

[1] School of Mathematics and Natural Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA
[2] Center of Excellence in Computational Chemistry, Department of Chemistry, Chulalongkorn University, Bangkok 10330, Thailand

* Correspondence: Email: ras.pandey@usm.edu; Tel: +6012664485.

Abstract: Unfolding of a coarse-grained COVN protein from its native configuration shows a linear response with increasing temperature followed by non-monotonic double peaks in its radius of gyration. The protein conforms to a random coil of folded segments in native state with increasingly tenuous and globular structures in specific temperature regimes where the effective dimensions of corresponding structures $D \approx 1.6$–2.4. Thermal agitation alone is not sufficient to fully eradicate its segmental folding as few local folds are found to persist around such residues as ^{65}L, ^{110}Y, ^{224}L, ^{374}P even at high temperatures.

Keywords: protein folding; nucleocapsid COVN protein; coarse-grain model; Monte Carlo simulation; thermal-induced folding

1. Introduction

CoVID-19 pandemic is attracting unprecedented attention [1–5] in investigating the corona virus and its constituents. Corona virus involves a number of proteins, RNA and a huge list of crowded inter- and intra-cellular constituents in its assembly and replication. In an initial investigation even with a coarse-grained computer simulation model it is not feasible to consider all constituents that are involved in its assembly and replication. We examine the structural dynamics of a nucleocapsid (COVN) protein [6] consisting of 422 residues which plays a critical role in packaging the viral genome RNA into ribonucleocapsid and virion assembly [7–9]. For the sake of simplicity and to develop a clear understanding of the basic nature of the conformational evolution, it would be interesting to examine the structural response of a free COVN as a function of temperature

before systematically including different types of proteins, solute, solvent etc. of the underlying host space.

2. Materials and method

'Protein folding' [10,11] remains an open problem despite enormous efforts for over half a century. Because of the enormity of challenges (e.g. time scale for huge degrees of freedom with all-atom approaches), coarse-graining [12–17] remains a viable choice to gain insight into the fundamental mechanism of conformational dynamics. Using a simplified yet efficient and effective coarse-grained model [18,19], a large-scale Monte Carlo simulation is performed to study the thermal response of COVN. Our coarse-grained model has already been used to investigate structural dynamics of such proteins as histones critical in assembly of chromatin [20], lysozyme [21] and alpha-synuclein [22] key in amyloid, protein (VP40) in ebola virus [23], membrane proteins [18,19] for selective transports, etc. COVN is represented by a chain of 422 residues in a specific sequence in a cubic lattice [18,19]. Each residue interacts with surrounding residues within a range (r_c) with a generalized Lennard-Jones potential,

$$U_{ij} = \left[\left| \varepsilon_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{12} + \varepsilon_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \right], \ r_{ij} < r_c \tag{1}$$

where r_{ij} is the distance between the residues at site i and j; $r_c = \sqrt{8}$ and $\sigma = 1$ in units of lattice constant. A knowledge-based [12–17] residue-residue contact matrix (based on a large ensemble of protein structures in PDB) is used as input for the potential strength ε_{ij} [14] in phenomenological interaction (1). With the implementation of excluded volume and limits on the covalent bond length constraints, each residue performs its stochastic movement with the Metropolis algorithm, i.e. with the Boltzmann probability $exp(-\Delta E/T)$ where ΔE is the change in energy between new and old position. Attempts to move each residue once defines unit Monte Carlo time step. All quantities are measured in arbitrary unit (i.e. spatial length in unit of lattice constant) including the temperature T which is in reduced units of the Boltzmann constant.

Simulations are performed on a 550^3 lattice for a sufficiently long time (10^7) steps with a number of independent samples (*100–1000*) over a wide range of temperatures. Different sample sizes are also used to verify the reliability of the qualitative trends from our data presented here. A number of local and global physical quantities such as radius of gyration, root mean square displacement of the center of mass, structure factor, contact map, etc. are examined as a function of temperature. The conformation of the protein exhibits a monotonous response from a random-coil of folded (globular) segments in native phase to tenuous fibrous conformations on raising the temperature; it exhibits a non-monotonic response with a re-entrant conformation involving enhanced globularity before reaching a steady-state conformation on further heating. While most segmental folds disappear in denatured phase while some persist even at a very high temperature (see below).

Before presenting our data, it is worth pointing out the justification of our model in context to investigation of proteins associated with the Corona virus which has only four structural proteins of which the envelope protein CoVE is the smallest with 76 residues. The primary and secondary structures of CoVE have shown to have three domains (see Figure 1 of Schoeman and Filelding [24]

and references therein) with N- and C-terminals separated by the transmembrane segment. These domains are faithfully identified and reproduced from the contact profiles [25] generated by the coarse-grained model used here. COVN is a relatively large protein as pointed above. Chang et al. [7] have identified N- (residues 45–181) and C-terminal (residues 248–365) domains of COVN that can bind to nucleic acids i.e. RNA. Thermal modulation of the contact profiles of COVN generated by the same coarse-grained model exhibits the evolution in segmental assembly that may be consistent with the responsiveness of the two regions (see below). Although it would be difficult to guaranty the results of a model for a quantitative comparison with laboratory observations, it appears that our coarse-grained model does capture some of the basic features of the proteins we have investigated so far.
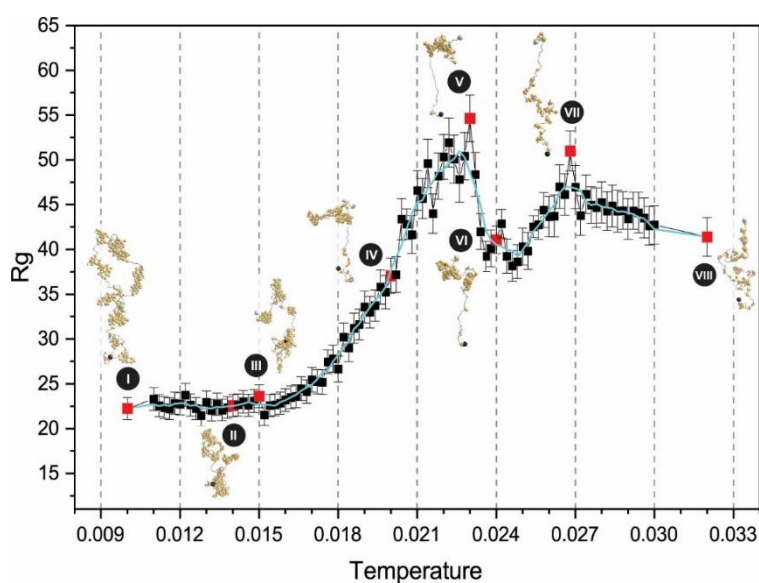
3. Results and discussion

Figure 1. Variation of the average radius of gyration (R_g) with the temperature. Some snapshots (at the time step t $= 10^7$) are included at representative temperatures: (*i*) *T=0.0100*, (*ii*) *T=0.0140*, (iii) *T=0.0150*, (iv) *T = 0.0200*, (v) *T = 0.0230* (first maximum), (vi) *T= 0.0240* (minimum), (vii) *T = 0.0268* (second maximum), (viii) *T= 0.0320*. Size of the self-organized segmental assembly represents the degree of globularization. In snapshots, gold spheres represent residues in contact, the large black sphere is the first residue ^{1}M and large grey sphere is the last ^{422}A (see Figure S1).

Figure 1 shows the variation of the average radius of gyration (R_g) with the temperature. At low temperatures (*T = 0.010–0.015*), the radius of gyration remains almost constant with its lowest magnitude ($R_g \sim 22.5$) in its native phase. Unlike many proteins (globular in native phase), COVN appears to be expanded into a random coil (see below) signature of an intrinsically disordered [7] protein. Raising the temperature (*T = 0.015–0.023*) leads to a monotonic increase to its maximum $Rg \sim 54.64 \pm 2.60$ at *T = 0.0230*. On further heating, the radius of gyration decreases sharply in a narrow range of temperature (*T = 0.023–0.025*) to a minimum value ($R_g \sim 38.17 \pm 1.72$) at *T = 0.0246* before it begins to increase with the temperature (*T = 0.0250–0.0268*) again until it reaches a second

maximum ($R_g \sim 51.00 \pm 2.24$) at $T = 0.0268$. Beyond the second peak, the radius of gyration continues to decay slowly towards its saturation with the temperature in denatured phase ($R_g \sim 41.4 \pm 2.17$ at $T = 0.032$, $R_g \sim 38.23 \pm 1.96$ at $T = 0.050$). Note that this trend is clear despite a relatively large fluctuation in data. To our knowledge, we are not aware of such a non-linear thermal response of such proteins. We believe this is due to unique structure of COVN.

Representative snapshots (Figure 1, see also Figure S1) of the protein at selected temperatures shows the variations in nature of the self-organizing structures over the range of temperature. For example, in native phase ($T = 0.010, 0.014$) we see local segmental folding with a chain of folded blobs in a random-coil-like conformation (see below) in contrast to a global folding one generally expects. Local folds begin to disappear at high temperatures but still persist in smaller sizes. Segmental folds appear to be distributed along the entire protein backbone at both maxima and at high temperatures in denature phase while the segmental folds at the minimum and in native phase are localized.
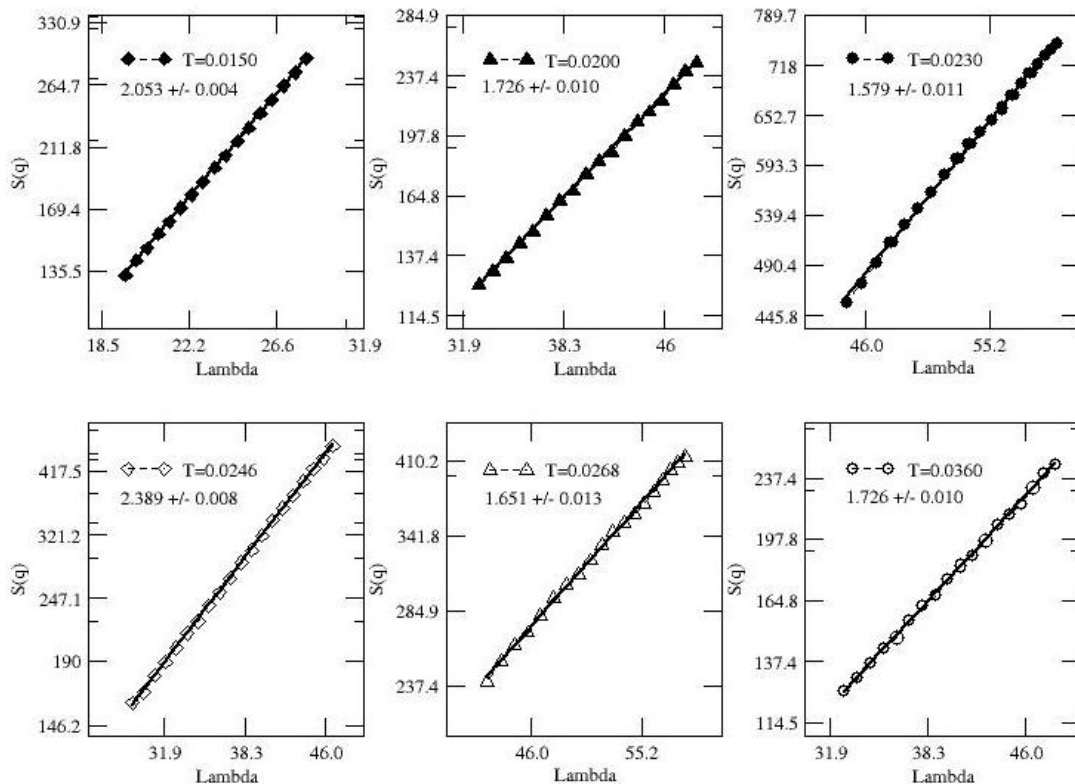

Figure 2. Structure factor $S(q)$ versus wavelength (lambda (λ)) comparable to radius of gyration of COVN on a log-log scale at representative temperatures.

How to quantify the distribution of residues over length scales ? To assess the mass (distribution), we have analyzed the structure factor $S(q)$ defined as,

$$S(q) = \langle \frac{1}{N} \left| \sum_{j=1}^{N} e^{-i\vec{q}\cdot r_j} \right|^2 \rangle_{|\vec{q}|} \tag{2}$$

where r_j is the position of each residue and $|q| = 2\pi/\lambda$ is the wave vector of wavelength λ. Using a power-law scaling $S(q) \propto q^{-1/\gamma}$, one may be able to evaluate the power-law exponent γ and estimate the spread of residues over the length scale λ. Overall size of the protein chain is described by its radius of gyration (R_g). Therefore, the structure factor over the length scale comparable to protein size ($\lambda \sim R_g$) can provide an estimate of the effective dimension D of the protein conformation via scaling the number of residues (N) $N \propto \lambda^D$ where $D = 1/\gamma$. Variations of $S(q)$ with the wavelength λ comparable to radius of gyration of the protein over the entire range of representative temperatures are presented in Figure 2.

In the native phase ($T = 0.0150$) where the radius of gyration is minimum ($Rg \sim 22.5$), the effective dimension $D \sim 2.053$ of the protein shows that the overall spread is not globular. It is rather random-coil, a chain of segmental globules (see Figure 1). In unfolding-transition regime ($T = 0.020$), the effective dimension $D \sim 1.726$ decreases while retaining its partial folding towards C-terminal (see below). Continuous increasing the temperature leads to maximum unfolding ($T = 0.0230$) where the protein chain stretches to its maximum gyration radius ($Rg \sim 55$) with lowest effective dimension $D \sim 1.579$ with a couple of unfolded segments (see below). Further heating leads to contraction with a lower radius of gyration ($Rg \sim 22.5$, $T = 0.0246$) with a higher effective dimension $D \sim 2.389$, which indicates more compact conformation than that in its native phase, a thermal-induced folding. The effective dimension begin to reduce with increasing the temperature further as the protein conformation approaches a tenuous structure, i.e. $D \sim 1.579$ at $T = 0.036$.
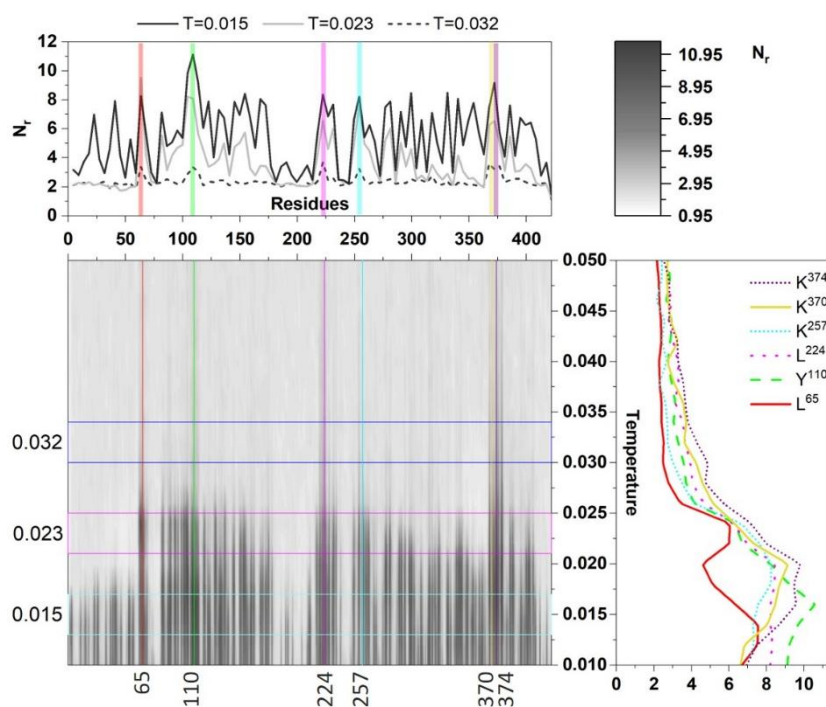

Figure 3. Average number (N_r) of residues in contact along the backbone of COVN as a function of temperature. Top Figure shows the contacts at representative temperatures in a native phase ($T = 0.015$), at the first (maximum) peak of the radius of gyration ($T = 0.0230$), and in a highly denatured phase ($T = 0.0320$). These regions of marked in the center three dimensional Figure with the scale at the upper right corner. Right Figure shows the thermal response of the contact profile of specific centers of folding.

Let us look closer into the local structures by examining the contact map in depth as presented in Figure 3 (see also Figure S2). First, we notice that the number of residues (N_r) within the range of interaction of each residues along the backbone, is higher at lower temperatures. However, the distribution of N_r is highly heterogenous and concentrated towards specific segments ([65]L, [110]Y, [224]L, [257]K, [370]K, [374]K). The degree of folds appears to be significant at these globularization centers (in particular segment [367]T-[380]A) even at higher temperatures although it is highest in native to denature transition region (see also Figure 1). In general, the modulation of the contact profiles shows the evolution in segmental assembly [Figure S2] that may be consistent with the responsiveness of N- and C- terminal domains [7]. Thermal response of contact profiles of each center of folding appears similar except [65]L which exhibits a non-linear (somewhat oscillatory) response (see the right section of Figure 3). However, it is worth pointing out the the response of the contact profile of [65]L resembles the thermal response of the radius of gyration. Despite the lowest magnitude of contacts (N_r) of [65]L with respect to other globularization centers i.e. [224]L, its unusual variations with the temperature (Figure 3) may induce global response in radius of gyration (Figure 1).

4. Conclusion

Thus, the thermal response of COVN protein is non-linear with a random coil of folded blobs in native phase to a systematic unfolding, refolding, and unfolding as the protein denatures on increasing the temperature. The radius of gyration increases on raising the temperature, first monotonically from a minimum in its native state to a maximum value. Further heating leads to a sharp decline (the protein contracts) in a narrow temperature range followed by increase (protein expands) again to a second maximum with a local minimum in between. The radius of gyration at the local minimum is larger than that in its native state but the segmental globularization is localized towards the second half (C-terminal) while the first half (N-terminal) of the protein acquire a fibrous configuration. Continued heating causes COVN to approach a steady-state value with a small contraction rate.

Scaling analysis of the structure factor is critical in quantifying the overall spread of COVN by evaluating its effective dimension D. In native phase, $D \sim 2.053$ ($T = 0.0150$, native phase), $D \sim 1.716$ ($T = 0.0200$, intermediate denature phase), $D \sim 1.579$ ($T = 0.0230$, first maximum), $D \sim 2.389$ ($T = 0.0246$, local minimum), $D \sim 1.651$ ($T = 0.0268$, second maximum), $D \sim 1.726$ ($T = 0.0360$, denatured). These estimated are consistent with the thermal response of the radius of gyration. Active zones of folded segments are identified from a detailed analysis of the contact map profile where the degree of folding can be quantified from the average contact measures. Segmental denaturing around residues such as [65]W, [110]Y, [224]L, and [374]P by technique other than thermal agitations may eradicate the specific functionality of COVN.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author contributions

All authors participated in discussion and agreed to investigate the protein folding of COVN. RBP proposed to use coarse-grain model in consultation with PS and WR. RBP and WR generated data. WR analyzed data. PS participated in results and discussion of the data.

References

1. Bzówka M, Mitusinska K, Raczynska A, et al. (2020) Molecular dynamics simulations indicate the COVID-19 mpro is not a viable target for small-molecule inhibitors design. *bioRxiv* doi.org/10.1101/2020.02.27.968008.
2. J Alsaadi EA, Jones IM (2019) Membrane binding proteins of coronaviruses. *Future Virol* 14: 275–286.
3. Wan Y, Shang J, Graham R, et al. (2020) Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J Virol* 94 e00127.
4. Xu J, Zhao S, Teng T, et al. (2020) Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* 12: 244.
5. Ahmed SF, Quadeer AA, McKay MR (2020) Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 12: 254.
6. UniProtKB-P59595(NCAP_SARS), 2020. Available from: https://www.uniprot.org/uniprot/P59595.
7. Chang CK, Hsu YL, Chang YH, et al. (2009) Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol* 83: 2255–2264.
8. Fang HJ, Chen YZ, Li MS, et al. (2009) Thermostability of the N-terminal RNA-binding domain of the SARS-CoV nucleocapsid protein: Experiments and numerical simulations. *Biophys J* 96: 1892–1901.
9. Chang C, Chen CMM, Chiang M, et al. (2013) Transient oligomerization of the SARS-CoV N protein–implication for virus ribonucleoprotein packaging. *PloS One* 8: e65045.
10. Sorokina I, Mushegian A (2018) Modeling protein folding in vivo. *Biol Direct* 13: 13.
11. Gershenson A, Gosavi S, Faccioli P, et al. (2020) Successes and challenges in simulating the folding of large proteins. *J Biol Chem* 295: 15–33.
12. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18: 534–552.
13. Miyazawa S, Jernigan RL (1996) Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 256: 623–644.

14. Betancourt MR, Thirumalai D (1999) 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.

15. Tanaka S, Scheraga HA (1976) Medium-and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* 9: 945–950.

16. Godzik A (1996) Knowledge-based potentials for protein folding: what can we learn from known protein structures ?. *Structure* 4: 363–366.

17. Huang SY, Zou X (2011) Statistical mechanics‐based method to extract atomic distance-dependent potentials from protein structures. *Proteins: Struct, Funct, Bioinf* 79: 2648–2661.

18. Kitjaruwankul S, Khrutto C, Sompornpisut P, et al. (2016) Asymmetry in structural response of inner and outer transmembrane segments of CorA protein by a coarse-grain model. *J Chem Phys* 145: 135101.

19. Boonamnaj P, Paudel SS, Jetsadawisut W, et al. (2019) Thermal-response of a protein (hHv1) by a coarse-grained MC and all-atom MD computer simulations. *Phys A* 527: 121310.

20. Fritsche M, Pandey RB, Farmer BL, et al. (2013) Variation in structure of a protein (H2AX) with knowledge-based interactions. *PLoS One* 8: e64507.

21. Pandey RB, Farmer BL, Gerstman BS (2015) Self-assembly dynamics for the transition of a globular aggregate to a fibril network of lysozyme proteins via a coarse-grained Monte Carlo simulation. *AIP Adv* 5: 092502.

22. Mirau P, Farmer BL, Pandey RB (2015) Structural variation of alpha-synuclein with temperature by a coarse-grained approach with knowledge-based interactions. *AIP Adv* 5: 092504.

23. Pokhrel R, Sompornpisut P, Chapagain P, et al. (2018) Domain rearrangement and denaturation in Ebola virus protein VP40. *AIP Adv* 8: 125129.

24. Schoeman D, Fielding BC (2019) Coronavirus envelope protein: current knowledge. *Virol J* 16: 69.

25. Pandey RB (2020) Thermal denaturation of a protein (CoVE) by a coarse-grained Monte Carlo simulation. arXiv preprint arXiv:2009.00049, 2020..