AIMS Biophysics, 9(4): 330–340.
DOI: 10.3934/biophy.2022027
Received: 26 July 2022
Revised: 01 November 2022
Accepted: 02 November 2022
Published: 09 November 2022

http://www.aimspress.com/journal/biophysics

Research article

Thermal denaturation of a coronavirus envelope (CoVE) protein by a coarse-grained Monte Carlo simulation

Panisak Boonamnaj[1], Pornthep Sompornpisut[1,*], Piyarat Nimmanpipug[2] and R.B. Pandey[3,*]

[1] Center of Excellence in Computational Chemistry, Department of Chemistry, Chulalongkorn University, Bangkok 10330, Thailand
[2] Department of Chemistry and Cluster of Excellence on Biodiversity based Economic and Society (B.BES-CMU), Chiang Mai University, Chiang Mai 50200, Thailand
[3] School of Mathematics and Natural Sciences, University of Southern Mississippi, Hattiesburg, MS 39406-5043, USA

*** Correspondence:** Email: pornthep.s@chula.ac.th, ras.pandey@usm.edu.

Abstract: Thermal response of an envelope protein conformation from coronavirus-2 (CoVE) is studied by a coarse-grained Monte Carlo simulation. Three distinct segments, the N-terminal, Trans-membrane, and C-terminal are verified from its specific contact profile. The radius of gyration (R_g) reveals a non-monotonic sub-universal thermal response: R_g decays substantially on heating in native phase under low-temperature regime in contrast to a continuous increase on further raising the temperature prior to its saturation to a random-coil in denature phase. The globularity index which is a measure of effective dimension of the protein, decreases as the protein denatures from a globular to a random-coil conformation.

Keywords: envelope protein; coronavirus; coarse-grained model; Monte Carlo simulation

1. Introduction

The outbreak of COVID 19 pandemic has caused an urgent need to understand a wide range of issues, both basic and applied, from epidemiology to diverse multi-scale molecular structures of its constitutive elements of the novel coronavirus (CoV-2) and its interaction with surroundings [1–4].

The RNA genome of CoV-2 is believed to encode only four major structural, sixteen non-structural and eleven accessory proteins. The four structural proteins including spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and the envelope (E) protein are key components for viral transmission and reproduction [4–9]. The CoVE protein is the smallest with 76 amino acid residues (^{1}M^{2}Y…^{76}V) (for sequence see https://www.uniprot.org/uniprot/P59637) [10] but plays an essential role in replication cycle of the virus such as formation of envelope as an integral membrane protein in viral assembly, budding and pathogenesis [4–16].

Very recently, Kuzmin et al. [17] have examined the conformational variability of CoVE protein in presence of a membrane environment using an all-atom MD along with a coarse-grained simulation. Because of 'prohibitively long time' with the atomistic simulation in 'exhaustive sampling of the conformational space' utility of coarse-grained approach implemented in this study is very well emphasized. They find that CoVE induces curvature in membrane and 'it has a preferable position' in already curved membranes with 'the C-termini localize to the convex region'. Apart from locating specific segments of the protein from the membrane sites, there is no estimate of the overall size of the protein or the underlying dynamics. For example, how the segments of protein perform its stochastic moves and how the overall dynamics (diffusive, non-diffusive, drift *etc.*) of the protein (i.e. that of its center of mass) sets in as its segments settles in the curved region. How its conformation and dynamics depends on temperature. It may be useful to explore the structure and dynamics of a protein monomer free from membrane as a function of temperature first. Then incorporate the environment such as membrane and other constitutive elements such as solvent to assess its effect on the structure and dynamics. We would like to focus on the folding dynamics of a free CoVE protein, a purely theoretical exercise as an idealized model system, the characteristic of which may be useful in hypothesizing its role in interpreting the laboratory data. Such a basic investigation has not been performed to our knowledge as presented here.

Sequence of CoVE protein and its specific characteristics such as formation of membrane ion channel, virulence intensity, *etc.* and other proteins are studied in depth [4–16]. Three distinct segments of the CoVE protein start with the hydrophilic amino (N) terminal with about 10 residues (^{1}M–^{10}G), followed by the hydrophobic transmembrane (TM) segment of about 25–27 residues (^{11}T–^{37}L), and ends with a relatively large hydrophilic carboxyl terminal consisting of the remaining residues (^{38}R–^{76}V). The structural details of how these segments interact, compete and cooperate is not fully known. The objective of this article is to investigate the effect of temperature on the structural dynamics by analyzing some of the local and global physical properties by a coarse-grained (CG) Monte Carlo simulation. The model and methods section is described in brief next followed by results and discussion and a conclusion.

2. Model and methods

2.1. Model

The model adopted here is based on an effective and efficient method involving bond-fluctuating mechanism on a discrete lattice with ample degrees of freedom to investigate complex issues in probing long time properties in polymer systems [18] that cannot be accessed by traditional MC or MD in a continuum space. This coarse-grained approach has been further developed over the years [19–21] with options to fine-grain and incorporate the specificity of amino acids to model proteins. The relevant

detail has been already described to examine structural dynamics of a number of proteins [19–21]. In coarse-grained description, we constructed a CG protein model of CoVE containing a chain of 76 amino acid residues [10] in a specific sequence on a discrete cubic lattice with ample degrees of freedom to describe essential motion and fluctuation for each residue in the protein. A residue is modeled using a single CG node with a cubic node of size $(2a)^3$ where a refers to the lattice constant. Consecutive nodes are tethered together by fluctuating covalent bonds of which the bond length between consecutive nodes varies between 2 and $\sqrt{(10)}$ in unit of lattice constant (a). A generalized Lennard-Jones potential (U_{ij}) is used to describe the interaction between each residue and surrounding residues within a range (r_c) as:

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

where r_{ij} is the distance between the amino acid residues at site i and j; $r_c = \sqrt{8}$ and $\sigma = 1$ in units of lattice constant. The potential strength ε_{ij} in phenomenological interaction (1) is based on a knowledge-based [22] residue-residue contact matrix which has been developed over many years [23–27] from a huge (and rapidly growing) ensemble of protein structures in Protein Data Bank (PDB).

2.2 Methods

The protein chain is placed in a random configuration initially since one can reach any microstate from any other in the vast landscape of statistical ensemble in thermodynamic equilibrium—a fundamental principle of statistical physics. Therefore, starting configuration is irrelevant. Each residue undergoes its stochastic movement with the Metropolis algorithm. The protein configurations are thus generated by movements of each residue with a probability proportional to the Boltzmann factor $exp(-\Delta E/T)$ where ΔE is the change in energy between new and old configurations of the underlying residues subject to strictly imposed excluded volume constraints. Monte Carlo time step is defined by attempts to move each residue once. Simulation parameters i.e. length in unit of lattice constant are measured in arbitrary unit and the temperature T in reduced units of the Boltzmann constant.

It is important to point out that connecting the arbitrary units with the real sample size and temperature i.e. Kelvin, and time scale (micro-second) quantitatively is difficult due to phenomenological nature of the residue-residue interaction and the coarse-grained representation. Calibration of the parameters used in simulation requires measurements of some common physical quantities in both computer simulations as well as laboratory experiments which is not feasible. However, based on the approximate size of residue and its representation and response of the physical quantities with respect to changes in temperature, one can estimate the order of magnitude.

All simulations reported here are performed on a 150^3 cubic lattice for a range of temperature $T = 0.010–0.030$ with 100 independent samples each with 10 million time steps; different sample sizes are also used to make sure that the qualitative results are independent of finite size effects. Both local and global physical quantities such as radius of gyration (R_g), root mean square displacement (RMSD) of the center of mass, structure factor (S), contact map, *etc.* are investigated as a function of reduced temperature. In the following 'temperature' refers to 'reduced temperature' and should not be confused with the absolute value.

3. Results and discussion

Snapshots of the protein (CoVE) at representative temperatures (T = 0.020, 0.024, 0.030) presented in Figure 1 illustrate the conformational changes as it denatures on raising the temperature. Obviously the protein conforms to a globular (compact) structure at low temperature (T = 0.020) and expands as it denatures (T = 0.030). The segmental globularity in the transmembrane segment of the protein is retained at least in part even at the high temperature (see below). Note that snapshots do not represent the average conformation; it is an illustrative mean that needs to be quantified. Therefore, some quantitative measures of segmental globularity are highly desirable. Average number (N_n) of residues within the range of interaction of each along the contour of the protein may provide some insight into the local restructuring as a function of temperature.

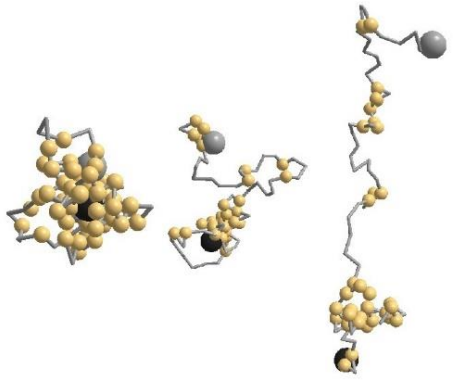

Figure 1. Snapshot of CoVE protein at the end of 10^7 MCS time. Large black and grey sphere represents the first residue (^{1}M) and the last residue (^{76}V) of the protein, respectively. Smaller golden spheres represent the presence of other residues within the range of interaction excluding the consecutively connected residues. Backbone grey lines are covalent bonds connecting the residues. Snapshot at the left (T = 0.020), center (T = 0.024), right (T = 0.030) are at the representative temperatures to illustrate denaturing of the protein.

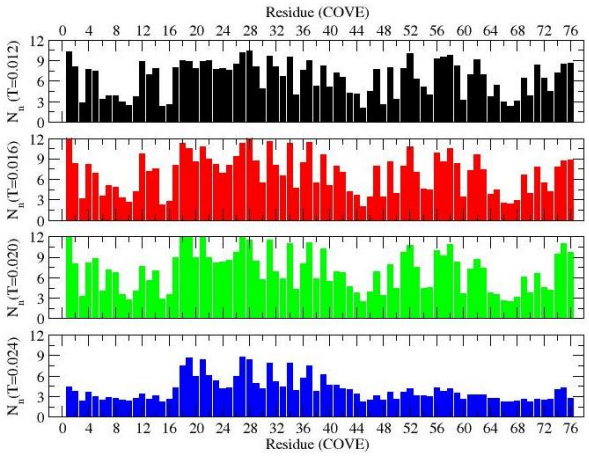

Figure 2. Residue contact profiles show average number (N_n) of surrounding residues around each residue of the protein within the range of interaction at a range of temperature (T = 0.012–0.024).

The contact profiles at representative temperatures are shown in Figure 2. At low temperatures (T = 0.012–0.020, in native phase), the coagulating centers are distributed sporadically particularly in N- and C-terminal regions along the backbone (e.g. ^{1}M^{2}Y, ^{12}L^{13}I^{14}V (in N-terminal), ^{51}L^{52}V^{53}K, ^{56}V^{57}Y^{58}V^{59}Y, ^{74}L^{75}L^{76}V (in C-terminus)) with relatively large N_n. The transmembrane segment (^{18}L–^{44}C), however, has the largest globular cluster with high number N_n of surrounding residues. Clearly, there is no significant change in contact profiles with the temperature in native phase. However, raising the temperature (T = 0.024) further leads to a substantial decrease in segmental globularity in both N- and C-terminals with low contact number N_n while retaining some globularity in TM segment with a relatively large N_n (see Figure 2). Thus, three distinct (N- and C-terminals separated by TM) segments can be identified easily from the evolution of contact profiles with the temperature.

The global structure and dynamics of the protein emerge from the stochastic movement of each residues and their configurational stabilities. Variation of the root mean square displacement of the center of mass (R_c) with the time step (t) at a range of temperature presented in Figure 3 show, how the asymptotic dynamics depends on the temperature. One can identify the type of dynamics by estimating the power-law exponent (v) by fitting the data with a scaling, $R_c \propto t^v$ where v = ½ represents diffusion. Accordingly, the protein is found to move extremely slowly at low temperatures with ultra-sub-diffusive dynamics (*i.e.* v = 0.05 at T = 0.020) and resumes diffusion at higher temperatures (*i.e.* v = 0.47 at T = 0.024, see Figure 3). Dynamics do affect the conformation of the protein which can be quantified by analyzing radius of gyration and the structure factor (see below).

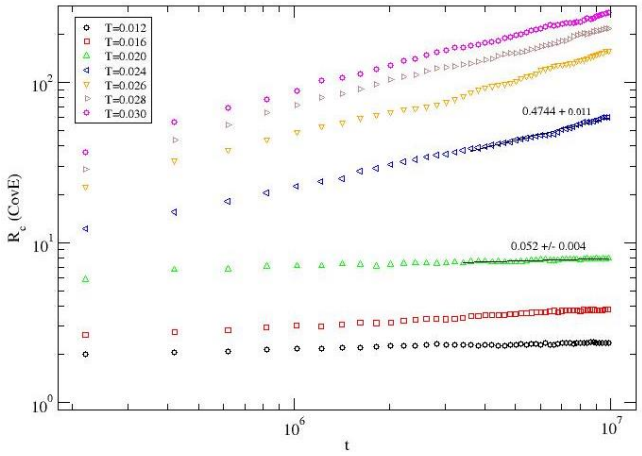

Figure 3. Average root mean square displacement (RMSD) of the center of mass (R_c) of the protein with the time step (t) versus temperature in range , T = 0.012–0.030 on a log-log scale. Slopes of the fitted data (solid black lines) in the asymptotic regime at a low (T = 0.012) and a high (T = 0.024) temperature are included with corresponding estimates.

How do we know that the protein chain has reached its steady-state equilibrium in 10 million time steps? It is important to check the convergence of physical quantities to make sure that system has reached steady-state within the range of simulation time steps in all such simulations. We have done here as in all our previous studies. For an illustration, variation of the radius of gyration with the time steps for a range of temperature is presented in Figure 4. We see that the radius of gyration has reached its steady-state equilibrium at almost all temperatures. It is not feasible to reach as good steady-state

at low temperatures as is the case in most such interacting systems and even in laboratories. However, the probability of reaching the same asymptotic magnitude of R_g at all temperatures (T below 0.020) in sub-native phase (i.e. T = 0.018, 0.016 in Figure 4) is extremely low. Since it is not feasible to extend this simulation for such a long asymptotic time regime, it is worth issuing a caution to very cautious readers. The data on R_g in this sub-native phase (T = 0.010–0.020) in the following Figures should be therefore considered with caution.

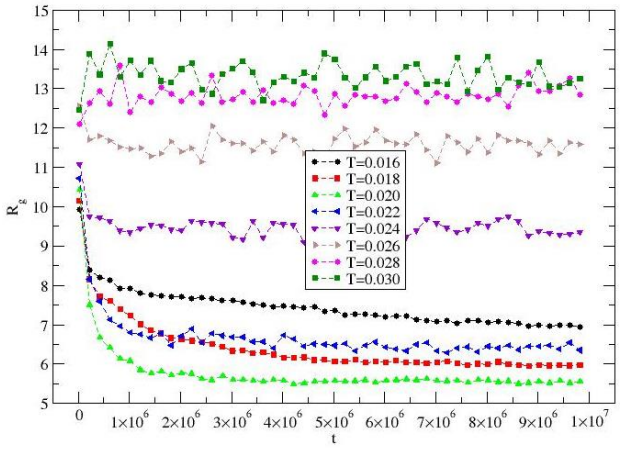

Figure 4. Variation of the average radius of gyration (R_g) of the protein with the time step (t) for a temperature range, T = 0.016–0.030.

The radius of gyration (R_g) of the protein is a measure of its average size. Variation of the average (over configurations from 5 million to 10 million time steps each with 100 independent samples) radius of gyration (R_g) with the temperature (T) is presented in Figure 5. In contrast to denature phase (T = 0.020–0.030) where the radius of gyration (R_g) increases on raising the temperature, it decays on increasing the temperature in the native phase (T = 0.010–0.020). Why the protein contract on raising the temperature in sub-native phase? At very low temperatures, segmental interactions dominate over thermal noise. On raising the temperature, the frozen or frustrated trapped segments begin to explore and equilibrate better and therefore result into a more compact conformation. This process continues until a certain temperature beyond which thermal noise take over the segmental interactions with the temperature and protein begins to expand. Such an opposite thermal response in native and denature phases is also observed in other membrane proteins in recent years [20]. Based on a crude calibration of the measurements of the physical quantities with the same coarse-grained approach and an all-atom MD simulation [20], we believe that the range of temperature T = 0.020–0.030 may lie in the range of T = 300–400 K roughly. In order to claim such thermal response, a universal or sub-universal characteristics for the membrane proteins, more studies on many different membrane proteins are needed. The thermal response of the nucleocapsid protein (CoVN) of COVID 19 is also non-monotonic [21] but very different from that of CoVE. We hope this results will stimulate further investigations in this direction.

It is worth pointing out in passing that the validation of this model is based on the response of protein at extreme temperature where residue-residue interaction becomes irrelevant and conformation of a protein chain behaves similar to that of an athermal polymer chain. That is the case here as in all our previous work based on this coarse-grained model. Conformational dynamics of a range of proteins

with various sizes are examined using this coarse-grained approach and observed different thermal response of proteins with comparable size. For example, proteins H3.1 and H2AX [19] are about the same size (136 and 140 residues) and exhibit very different thermal response to temperature: on raising the temperature, R_g of H3.1 increases continuously before saturation while R_g of H2AX increases to a maximum value followed by decay (a non-monotonic response different from the one observed here for CoVE). Thus, the specificity of residues and its distribution (i.e. sequence) in a protein pays a critical role in their collective response. Unfortunately, we do not have a data on a protein of the COVID related protein with the size comparable to CoVE to compare here. However, based on the analysis on many proteins (some cited in this study) with this approach, the conclusion on the thermal response of CoVE is reliable. A reader can always check it out by carry out simulation independently on an idealized model like the one presented here.

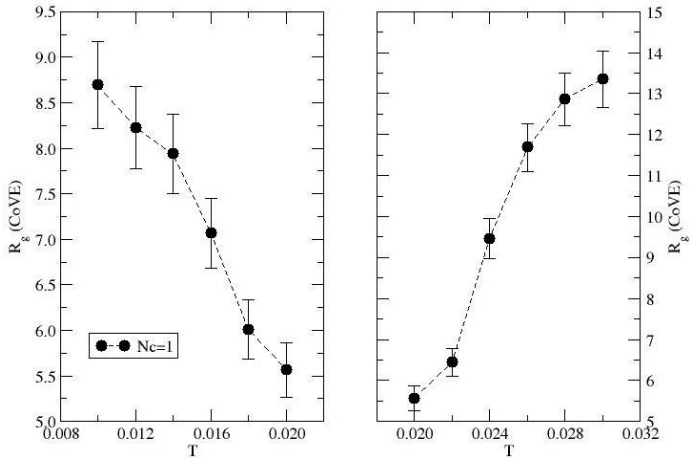

Figure 5. Average radius of gyration (R_g) of CoVE versus temperature.

Analysis of the structure factor can also provide insight into the global conformation of the protein. The structure factor ($S(q)$) is defined as:

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

where r_j is the position of each residue. The magnitude of the wave vector (q) of wavelength λ is given by $|q| = 2\pi/\lambda$. Using a power-law scaling for the structure factor with the wave vector, $S(q) \propto q^{-1/\gamma}$, one can evaluate the exponent γ. Variations of the structure factor with the wavelength (comparable to radius of gyration) is presented in Figure 5 for representative temperatures.

In general, the radius of gyration (R_g) of a polymer chain shows a power-law scaling with number (N) of monomer (residues), $R_g \propto N^x$ with a well-defined exponent x. Since a protein is a hetero-polymer chain of residues, one may use the same scaling law here for the protein with the radius of gyration comparable to wavelength ($R_g \sim \lambda$). Then the power-law exponent $x = \gamma$ and one may be able to estimate the globularity index g which is the effective dimension (D) for the spread of the protein ($g \cong D$). Since, $N \propto R_g^D$, $D = 1/\gamma$. Higher magnitude of D represents more compact structure with higher degree of globularity. Note that this scaling analysis is an estimate and should not be taken as a precise measurement. Obviously D cannot be higher than 3, however high value indicates more compact conformations. Clearly, the protein remains globular (three dimensional) at low temperatures in native

phase and it denatures into a ramified random coil ($D \sim 2$) structures at high temperatures (see Figure 6). It is important to point out that the such a vast changes in conformation of the protein in sub-native i.e. low temperature (T = 0.010–0.020) to native-to-denature phase i.e. at higher temperature regime (T = 0.020–0.030) may pose the possibility of COVE to be an intrinsically disordered protein at least in some range of temperature and time scales. For example, the segmental interactions in the native phase may be somewhat similar to interaction of a protein with its underlying environment e.g. a specific surface [28] even in the absence of explicit external surface.

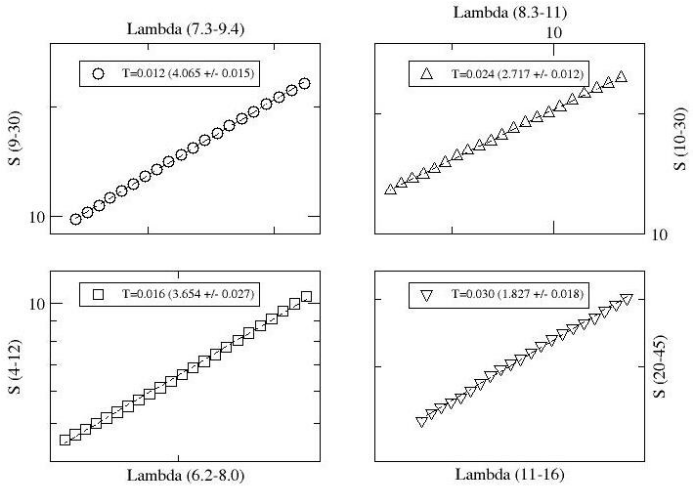

Figure 6. Structure factor (S) of the protein versus wave length (lambda) at representative low and high temperatures. Data points are selected with wavelength comparable to radius of gyration of the protein at corresponding temperatures with the estimates of slopes. The number along the axes are the range of the axes (starting and end points).

4. Conclusions

Monte Carlo simulations are performed with a coarse-grained model of envelope membrane protein CoVE of the human coronavirus to assess its structural variability as a function of temperature. The evolution of the protein's conformation as a function of temperature is clearly seen in visualization via snapshots and animations. Segmental organizing of residues is quantified by examining local physical quantities such as average contact and mobility profiles as a function of temperature. From the variations in contact profiles as a function of temperature, it is feasible to identify three distinct segments (N-terminal and C-terminal separated by a transmembrane segment) which is consistent with the previous studies on CoVE [4,5,11–16]. The radius of gyration exhibits a non-monotonic thermal response: the protein CoVE contracts on heating in native phase and expands continuously on raising the temperature before reaching a saturation in denatured phase. Such a unique non-monotonic thermal response has been recently observed in other membrane proteins (unrelated to coronavirus). This may lead to speculate that thermal response may be a unique characteristic to identify different classes of proteins—obviously more studies are needed to confirm this observation. Thermal response of CoVE is different from another structural protein CoVN (a nucleocapsid protein) of the coronavirus [21]. Variation of the structure factor with the wave vector is studied in detail for a wide range of temperature. Scaling analysis of the structure factor provides a quantitative estimate of the overall globularity with

an effective dimension (*D*). CoVE is found to remain globular ($D \sim 3$) in its native phase and unfolds continuously to a random coil ($D \sim 2$) conformation in denature phase on raising the temperature.

Acknowledgments

The authors acknowledge HPC at The University of Southern Mississippi supported by the National Science Foundation under the Major Research Instrumentation (MRI) program via Grant # ACI 1626217.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Panisak Boonamnaj: Conceptualization, Pornthep Sompornpisut: Conceptualization, Writing-Review & Editing, Administration, Piyarat Nimmanpipug: Conceptualization, R.B. Pandey: Conceptualization, Methodology, Investigation, Formal Analysis, Writing-Original Draft, Review & Editing.

References

1. Giri R, Bhardwaj T, Shegane M, et al. (2021) Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell Mol Life Sci* 78: 1655–1688. https://doi.org/10.1007/s00018-020-03603-x
2. 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: e00127-20. https://doi.org/10.1128/JVI.00127-20
3. 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. https://doi.org/10.3390/v12020244
4. Schoeman D, Fielding BC (2019) Coronavirus envelope protein: current knowledge. *Virol J* 16: 69. https://doi.org/10.1186/s12985-019-1182-0
5. Siu YL, Teoh KT, Lo J, et al. (2008) The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J Virol* 82: 11318–11330. https://doi.org/10.1128/JVI.01052-08
6. Pawłowski PH (2021) Charged amino acids may promote coronavirus SARS-CoV-2 fusion with the host cell. *AIMS Biophys* 8: 111–120. https://doi.org/10.3934/biophy.2021008
7. Pawłowski PH (2022) SARS-CoV-2 variant Omicron (B. 1.1. 529) is in a rising trend of mutations increasing the positive electric charge in crucial regions of the spike protein S. *Acta Biochim Pol* 69: 263–264. https://doi.org/10.18388/abp.2020_6072
8. Wu C, Yin W, Jiang Y, et al. (2022) Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19. *Acta Pharmacol Sin*: 1–13. https://doi.org/10.1038/s41401-021-00851-w

9. Naqvi AAT, Fatima K, Mohammad T, et al. (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. *BBA-Mol Basis Dis* 1866: 165878. https://doi.org/10.1016/j.bbadis.2020.165878
10. Consortium UniProt (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49: D480–D489. https://doi.org/10.1093/nar/gkaa1100
11. Wilson L, Mckinlay C, Gage P, et al. (2004) SARS coronavirus E protein forms cation-selective ion channels. *Virology* 330: 322–331. https://doi.org/10.1016/j.virol.2004.09.033
12. Torres J, Wang J, Parthasarathy K, et al. (2005) The transmembrane oligomers of coronavirus protein E. *Biophys J* 88: 1283–1290. https://doi.org/10.1529/biophysj.104.051730
13. Liu J, Sun Y, Qi J, et al. (2010) The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J Infect Dis* 202: 1171–1180. https://doi.org/10.1086/656315
14. Venkatagopalan P, Daskalova SM, Lopez LA, et al. (2015) Coronavirus envelope(E) protein remains at the site of assembly. *Virology* 478: 75–85. https://doi.org/10.1016/j.virol.2015.02.005
15. Surya W, Li Y, Torres J (2018) Structural model of the SARS coronavirus E channel in LMPG micelles. *BBA-Biomembranes* 1860: 1309–1317. https://doi.org/10.1016/j.bbamem.2018.02.017
16. Gupta MK, Vemula S, Donde R, et al. (2021) *In-silico* approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus envelope protein ion channel. *J Biomol Struct Dyn* 39: 2617–2627. https://doi.org/10.1080/07391102.2020.1751300
17. Kuzmin A, Orekhov P, Astashkin R, et al. (2022) Structure and dynamics of the SARS-CoV-2 envelope protein monomer. *Proteins* 90: 1102–1114. https://doi.org/10.1002/prot.26317
18. Binder K (1995) *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, 1 Eds., New York: Oxford University Press.
19. Fritsche M, Pandey RB, Farmer BL, et al. (2013) Variation in structure of a protein (H2AX) with knowledge-based interactions. *PLoS One* 8: e64507. https://doi.org/10.1371/journal.pone.0064507
20. 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. *Physica A* 527: 121310. https://doi.org/10.1016/j.physa.2019.121310
21. Rangubpit W, Sompornpisut P, Pandey RB (2021) Thermal-induced unfolding-refolding of a nucleocapsid COVN protein. *AIMS Biophys* 8: 103–110. https://doi.org/10.3934/biophy.2021007
22. 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. https://doi.org/10.1110/ps.8.2.361
23. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18: 534–552. https://doi.org/10.1021/MA00145A039
24. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and treading. *J Mol Biol* 256: 623–644. https://doi.org/10.1006/JMBI.1996.0114
25. 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. https://doi.org/10.1021/ma60054a013

26. Godzik A (1996) Knowledge-based potentials for protein folding: what can we learn from known protein structures? *Structure* 4: 363–366. https://doi.org/10.1016/s0969-2126(96)00041-x
27. Huang SY, Zou X (2011) Statistical mechanics-based method to extract atomic distance-dependent potentials from protein structures. *Proteins* 79: 2648–2661. https://doi.org/10.1002/prot.23086
28. Caetano DLZ, Metzler R, Cherstvy AG, et al. (2021) Adsorption of Lysozyme into a charged confining pore. https://doi.org/10.1101/2021.07.11.451934

AIMS Press