Long-Time Simulation of Temperature-Varying Conformations of SARS-CoV-2 Spike Glycoprotein on IBM Supercomputers

Meichen Song
Applied Mathematics Department
Stony Brook University
Stony Brook, NY, USA
meichen.song@stonybrook.edu

Peng Zhang
Applied Mathematics Department
Stony Brook University
Stony Brook, NY, USA
peng.zhang@stonybrook.edu

Changnian Han
Applied Mathematics Department
Stony Brook University
Stony Brook, NY, USA
changnian.han@stonybrook.edu

Ziji Zhang
Applied Mathematics Department
Stony Brook University
Stony Brook, NY, USA
ziji.zhang@stonybrook.edu

Yuefan Deng
Applied Mathematics Department
Stony Brook University
Stony Brook, NY, USA
yuefan.deng@stonybrook.edu

***Abstract*-- We investigated the conformational variations and phase transition properties of the spike glycoprotein (S-protein) of the coronavirus SARS-CoV-2 at temperatures ranging from 3℃ to 300℃ on the IBM Power9-based AI supercomputers. Our microsecond time-scale molecular dynamics simulations on the temperature-dependent properties of the S-protein help reveal (1) the protein is structurally stable at 3℃, consistent with observations that the virus stays active for more than two weeks in the cold supply chain; (2) the protein's side chains are more oscillatory from 60℃ to 80; and (3) the protein's survival time at temperatures 3℃, 20℃-37℃, 60℃-70℃, 80℃-95℃, and 300℃ groups roughly to week-, day-, minute- and microsecond-scales, respectively, by correlating with in vitro data.**

Keywords—Protein, molecular dynamics, conformation.

I. INTRODUCTION

The spread of SARS-CoV-2 engulfs the world from tropical countries like India [1] at 51℃ or higher to cold supply chain [2] at 3-4℃. Similar to other coronaviruses, the spike glycoprotein (S-protein) outer membrane [3], interacts with host cell targets such as ACE2, CD26, Ezrin, cyclophilins, and other cell adhesion factors and such interactions are important for cell adhesion and virulence, and thereby causing infection. Its conformation and binding sites are understood while its other properties including infection intensity and dynamics under such ambient conditions as temperatures are still elusive.

Ambient temperature may cause conformational change, and its functions such as infectivity, of the virus. These impacts were evidenced in experiment studies including electron microscopy that show conformational change at 37℃ and pH 8.0 in the coronavirus S-protein and the importance of these changes is evident in the variations of virus infectivity [4]. The past pandemics showed a strong correlation of the extent of contagion to a temperature similar to that of the current still ongoing coronavirus pandemic [5] . Experiments revealed that the infectivity of the influenza virus is preserved depending on the temperature of the aerosol containing viral particles. COVID-19 and influenza have a similar disease presentation. While the influenza has been studied for decades, the properties of COVID-19 will take longer to fully understand, and the data available for understanding is still sparse. The challenges are that as with the most highly contagious viruses, it is too difficult and too dangerous to handle them in the most traditional laboratories, not to mention exhaustive tests, necessitating the computational simulations. However, simulating large proteins such as viruses with hundreds to thousands of residues is extremely challenging even for very fast supercomputers, due to the need to simulate a large system for long time scales. Until now, there is no known modeling of the SARS-CoV-2 at long-time scales. **Our contribution:** We conduct the microsecond time scales, likely the longest, simulations of the S-protein on the latest IBM supercomputers and reveal the S-protein's conformational changes at various eventful temperatures to bridge the data gaps from in vitro experiments to viral transmission assessment. HPC help seek the critical temperature at which the S-protein may transit to another phase in which it depletes its infectivity and even denature. This work highlights the benefits of HPC for enabling the quantitative first-principle's studies of the SARS-CoV-2 structures and complement the sparse and inconsistent in vitro experiments with high-infectious virus cultivation.

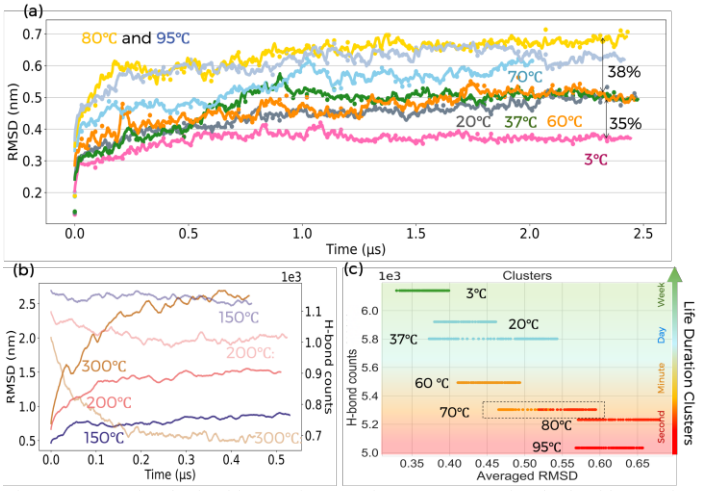

Fig. 1. RMSD for the backbone of S-protein **(a)** ; RMSD for the backbone of S-protein (represented as darker lines) and mainchain-mainchain H-bonds (represented as lighter lines). **(b)**; temperatures grouped by K-means in terms of their life duration based on features captured from RMSD and # of water-protein H-bonds **(c).**

II. HPC Experiments and Performance

Our MD simulations of S-protein (PDB file 6VXX) immersed in a H_2O box are performed at eventful or purposeful temperatures: 3℃ (of a cold supply chain), 20℃ (typical room temperature), 37℃ (normal human body temperature), 60℃, 70℃, 80℃, 95℃ (temperatures selected to locate the critical temperature, if exist), 150℃, 200℃ and 300℃ (to detect denaturation phases), using the CHARMM27 force field under NVT ensemble with a time step size of 2.5 fs. The total number of atoms is 805,218, in which 45,156 (5.6%) for the protein and 760,047 (94.4%) for water. In each study, 6.5-ns of energy minimization is always performed by gradient optimization and the energy-minimized structure is served as the initial state for MD simulation. For the production runs, additional 0.2-ns equilibration runs are carried out to drive the system to a desired temperature. The raw trajectory obtained from a simulation using a periodic boundary condition was processed by re-centering the protein in the center of the box in each frame. Furthermore, each protein frame was then aligned, such that the protein was superimposed on the reference structure by overall protein rotation and translation. Last, the processed trajectory of atomic positions was used for the RMSD (root-mean-square deviation) of backbone and Hbonds counts to measure the S-protein's thermal inabilities.

The in silico data are applied to cluster the temperatures to generate a rough grouping of the protein's survival duration from weeks to seconds. The samples are chosen from 0.3-2 μs simulation data with 100ns window size and 10 ns stride. The sample features consist of RMSD's average, standard divation, kurtosis, skewness and the # of water-protein Hbonds' average and variation which are taken as constants for each temperature. After normalized the sample features, the K-means are applied to cluster samples for all temperatures under 100℃.

Our experiments are conducted on the AiMOS supercomputer, a heterogeneous system that includes IBM Power9 CPUs and Nvidia V100 GPUs. To compare speed, we tested the same program on a local SeaWulf cluster (Intel Xeon E5 CPUs). The below figure compared the speedup via the nodes, showing a speedup of **7x** for AiMOS over SeaWulf. That is, AiMOS can reduce the 1-year task of a cluster to 2 months, thus enabled microsecond-scale MD simulation for a long-time experiment (Fig. 2).

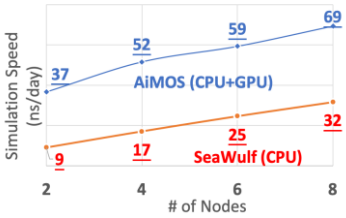

Fig. 2. Speeds on AiMOS and SeaWulf

III. Results and Analysis

(1) S-protein **experiences a phase transition in temperatures 60°C~80°C:** Fig. 1. (a) showed the RMSD increases by 38% as the temperature from 60℃ to 80℃, signaling a critical heat denaturation in this range.

(2) S-protein **stays stable at cold temperature:** By comparing results for 3°C and 20°C in Fig. 1. (a), the RMSD at 3℃ is 35% lower than that at 20℃, matching in-vitro experiments where salmon-attached corona-virus at 4 ℃ remains infectious for more than one week [2] with insignificant reduction of infectious titer on Day-14 [6].

(3) S-protein is **high temperature resistent:** although the # of mainchain-mainchain Hbonds reduces by about 30% within 0.5 μs, the most α-helix still remain stable (Fig. 1. (b)).

(4) Grouped temperature ranges by K-means in terms of the virus' survival time based of features captured from RMSD and # of water-protein H-bonds agree with in vitro data [6]: 3 ℃, 20 ℃ -37 ℃, 60 ℃ -70 ℃ and 80 ℃ -95 ℃ groups correspond to week-, day-, minute- and second-scale, respectively (Fig. 1. (c)).

IV. Conclusions and Future Work

MD simulation enabled by the world's top supercomputers including Anton 2 [7], Summit [8], and RIKEN [9] has been the primary tool for analyzing large molecules for drug designs and conformational analysis. For the first time, we present the microsecond-scale MD studies of the temperature-varying conformation of a life-threatening S-protein. Our simulations, helping understand the thermal properties of the S-protein, corroborate well with the in-vitro experiments and, conversely, guide additional in-vitro experiments. Our continued efforts will analyze more of the S-protein's properties including the Gibbs free energy, H-bonds, SASA, and gyration angles for analzying structures and, possibly, functions.

Acknowledgment

The project is sponsored by the OVPR&IEDM COVID-19 Seed Grant, PIs: P. Zhang, Y. Deng, M. Rafailovich, M. Simon. The project is supported by the SUNY-IBM Consortium Award, PI: Y. Deng, Co-I: P. Zhang. Core simulations were conducted on the AiMOS supercomputer at RPI and the WSC Cluster at IBM through an IBM Faculty Award FP0002468 (PI: Y. Deng).

References

[1] K. Sarkar, S. Khajanchi, and J. J. Nieto, "Modeling and forecasting the COVID-19 pandemic in India," Chaos, Solitons & Fractals, vol. 139, p. 110049, 2020.

[2] M. Dai, H. Li, N. Yan, J. Huang, L. Zhao, S. Xu, et al., "Long-term survival of salmon-attached SARS-CoV-2 at 4° C as a potential source of transmission in seafood markets," bioRxiv, 2020.

[3] W. Song, M. Gui, X. Wang, and Y. Xiang, "Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2," PLoS pathogens, vol. 14, no. 8, p. e1007236, 2018.

[4] A. Krowiak, "Symbolic computing in spline-based differential quadrature method," *Comm. Numer. Meth. Eng.,* vol. 22, pp. 1097-1107, 2006.

[5] J. Steel *et al.*, "Transmission of a 2009 pandemic influenza virus shows a sensitivity to temperature and humidity similar to that of an H3N2 seasonal strain," *J. Virol.,* vol. 85, pp. 1400-2, Feb 2011

[6] A. W. Chin et al., "Stability of SARS-CoV-2 in different environmental conditions," 2020.

[7] D. E. Shaw *et al.*, "Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer," in *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2014, pp. 41-53.

[8] B. Acun *et al.*, "Scalable molecular dynamics with NAMD on the Summit system," *IBM journal of research and development,* vol. 62, pp. 4: 1-4: 9, 2018.

[9] T. Komatsu *et al.*, "COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease," *Mendeley Data,* vol. 10, 2020.