Modeling of the Thermal Properties of SARS-CoV-2 S-protein

Meichen Song

SUNY Stony Brook: Stony Brook University <https://orcid.org/0000-0001-8054-0602>

Ziyuan Niu

Stony Brook University

Karin Hasegawa

Stony Brook University

Yuefan Deng

Stony Brook University

Miriam Rafailovich

Stony Brook University

Marcia Simon

Stony Brook University

Peng Zhang (✉ peng.zhang@stonybrook.edu)

Stony Brook University

Research Article

Keywords: SARS-CoV-2, S-protein, Molecular dynamics, Thermodynamics, Conformational state

Posted Date: August 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-747815/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Modeling of the Thermal Properties of SARS-CoV-2 S-protein

Meichen Song¹, Ziyuan Niu¹, Karin Hasegawa¹, Yuefan Deng¹, Miriam Rafailovich², Marcia Simon³, Peng Zhang¹

Departments of ¹Applied Mathematics and Statistics, ²Materials Science and Chemical Engineering, ³Oral Biology and Pathology, Stony Brook University, NY 11794, United States.

⁴Email: peng.zhang@stonybrook.edu

Abstract

We present the thermal and conformational states of the spike glycoprotein (S-protein) of SARS-CoV-2 at six temperatures ranging from 3°C to 95°C by all-atom molecular dynamics (MD) μ s-scale simulations. While corroborating with clinical results of the temperature impact on the COVID-19 infection, we examine the potential phase transitions of the S-protein in the temperature range and our simulation results revealed the following thermal properties of the S-protein: (1) It is structurally stable at 3°C, agreeing with observations that the virus stays active for more than two weeks in the cold supply chain; (2) Its structure varies more significantly for temperature window of 60°C to 80°C than in all other windows; (3) The sharpest structural variations occur near 60°C, signaling a plausible critical temperature nearby; (4) The maximum deviation of the receptor-binding domain at 37°C suggests the anecdotal observation that the virus is most infective at 37°C; (5) The *in silico* data agree with reported experiments of the SARS-CoV-2 survival times from weeks to seconds by our clustering approach analysis.

Keywords SARS-CoV-2 · S-protein · Molecular dynamics · Thermodynamics · Conformational state.

Introduction

The SARS-CoV-2 engulfs the world even for tropical countries, such as India [1] and Indonesia with the highest temperature of 56°C, and also it remains infectious for weeks at the low temperatures around 3°C-4°C for the cold supply chain, such as salmon-attached SARS-CoV-2 [2, 3]. Thus, temperature-related infection issues arouse significant attention, resulting in the well-documented recent studies [4-6] of the correlation between the spread of the COVID-19 and climate. Research in Indonesia on the correlation of weather conditions including temperature, humidity and precipitation with the COVID-19 pandemic shows that only average temperature is significantly correlated [6]. Another study, which included all countries impacted by COVID-19, confirmed the link between climatic factors, particularly a strong link with temperature, and discovered a negative correlation between average temperature and the frequency of SARS-CoV-2 illnesses [5]. Research examining the relationships of factors at the country level is also important for correlating temperature with the instantaneous reproduction number R_t of SARS-CoV-2: The R_t decreases to a minimum with the temperature approaching 11°C, increases from 11°C to 20°C, then decreases at temperatures above 20°C [4]. Such impact of temperature on the infectivity and the transmission of SARS-CoV-2 has been confirmed by the factor analysis.

The effect of temperature was also verified in biological experiments. At the virus level, the ambient temperature may cause changes in the virus structures and functions such as infectivity. These impacts were evident in experimental studies, showing that coronaviruses are thermolabile, for example, SARS-CoV is inactivated after 75°C heat for 15 mins, and MERS becomes inactivated after 65°C heat for 1 min [7, 8]. The structural alterations in spike glycoprotein (S-protein), which is critical for cell adherence and pathogenicity [9, 10], at 37°C and pH 8.0 are crucial in understanding viral infectivity [11]. These impacts were not only apparent in the previous experimental studies for SARS-CoV and MERS, but also in the recent laboratory studies about SARS-CoV-2 [2, 12-17]. In the highly publicized reports, the SARS-CoV-2 has shown high sensitivity to heat, for example, the virus attached to salmon remains viable at 4°C for 8 days and at 25°C for 2 days [2] and the virus remains infectious in aerosols at 21°C-23°C and 40% relative humidity for 3~24 hours [18]. When the incubation temperature reaches 70°C, the time for virus inactivation was reduced to 5 mins [14]. The other laboratory experiments also verify the same observations that SARS-CoV-2 was found to be inactive after 60°C heat for 15 mins but remain active at 37°C for at least 24 hours [17]. These confirm that SARS-CoV-2 gradually loses viability with increasing temperature. While

the data available for understanding is still sparse and handling these clinical specimens poses a biosafety risk to both healthcare professionals and laboratory workers [17], necessitating the computational simulations.

Understanding the biochemical and thermolabile properties of SARS-CoV-2 requires a comprehensive understanding of the structure at the atomic level. The S-protein is one of the coronavirus's initial and largest structural proteins [19]. The outer membrane S-protein, which is similar in the genome to SARS-CoV [20], is the primary host interaction protein with host cell targets such as ACE2 and is critical for cell adherence and pathogenicity [9, 10]. The temperature dependence of the SARS-CoV-2 characteristics, notably for S-protein, is interesting in medical prediction because ambient temperatures first alter the structure of the virus's protein membrane. The S-protein has two domains: the S1 domain is on top and comprises the portion that interacts directly with host cell receptors; the S2 domain, in contrast to the S1 subunit, forms the stalk of the S-protein [21], facilitating virion fusion with cellular membranes [22]. The receptor binding domain (RBD) is responsible for binding to ACE2, which is the initial step for entry into target cells [23]. *In vitro* binding studies confirmed that the RBD on SARS-CoV-2 binds to ACE2 with a low nano band affinity, indicating that it is a crucial functional component in the S1 domain that induces the binding to ACE2. [22, 24]. The RBD of S1 domain undergoes hinge-like conformational changes, transitioning from a closed state that hides the determinants for receptor binding causing inaccessible to the receptor to an open state that reveals the determinants on RBD so that becomes accessible to the receptor [10]. At the outermost of the protein, the N-terminal domain (NTD) is increasingly exposed. The NTD and RBD sections are more versatile, allowing for structural adaptation to host receptors [25]. The structure and binding sites of such S-proteins are well established, while other characteristics especially the infection intensity and stability under external factors such as temperature remain elusive.

The use of all-atom molecular dynamics (MD) simulation is critical for gaining a deeper understanding of its characteristics at the atomic level. MD modeling has been used to investigate the binding characteristics of SARS-CoV-2. Long-time MD simulations for relatively smaller systems have been used to effectively signal the binding process of the SARS-CoV-2 main protease (Mpro, 3CL hydrolase) with seven HIV inhibitors [26], while short-time MD simulations for larger systems have been used to reveal the properties of SARS-CoV-2 binding ACE2 [27]. By free energy analysis, SARS-CoV-2 may withstand a higher temperature than SARS-CoV and RBD-ACE2 binding for SARS-CoV-2 is more temperature-sensitive than SARS-CoV [27]. MD simulations

also demonstrate a temperature-dependent binding affinity of SARS-CoV-2 to the ACE2 receptor [28]. SARS-CoV-2-ACE2 binding was less affinitive and stable at 40°C (18nM) than at 37°C (6.2nM), while SARS-CoV-ACE2 binding was not (6.4nM vs. 8.5nM), according to the study [28]. While such binding properties of the S-protein have been thoroughly investigated, long-time scale MD simulations to study the temperature influence on the S-protein's conformation are uncommon. 200ns MD simulations at 10°C-50°C provided some preliminary structural states at these temperatures [29]. Due to the requirement to simulate over extended time scales, even very powerful supercomputers struggle to simulate large proteins like viruses.

To explore the further effect of the more eventful temperatures on the conformational change of the S-protein, we conduct microsecond scale simulations at 3°C (of a cold supply chain), 20°C (typical room temperature), 37°C (normal human body temperature), 60°C, 70°C, 80°C, 95°C (temperatures selected to locate the critical temperature, if exist). In addition to the structure states, the *in vitro* data are correlated to relate the life scale of the virus to the conformational change of the S-protein. The simulations are long enough to collect conformational samples of the S-protein for applying statistical analysis to study the effect of temperature on the virus' stability and infectivity. The analysis is based on a variety of measurements, including root-mean-square deviation (RMSD), the number of hydrogen bonds (H-bonds), solvent-accessible surface area (SASA) of the whole protein and the deviation of the RBD, from different perspectives and the consistency of all measurement is verified. Additionally, the analysis at residue level based on root-mean-square fluctuation (RMSF) and newly introduced structural divergence are performed to isolate the residues that are more temperature-sensitive and to assess the overall structural divergence at a given temperature.

Our simulations of μ s-long atomic resolutions have demonstrated the S-protein's conformational changes at seven well-chosen temperatures with which we discovered the existence, and the possible critical temperature, of a phase transition that rives the states of infectivity and non-infectivity. The temperatures are clustered by unsupervised learning based on a combination of *in silico* measurements, and the clustering result agrees well with published *in vitro* data. Finally, the spike opening, represented by the deviation of the RBD [30], at different temperatures are compared. The simulations in this work are conducted on the most powerful IBM supercomputers to help achieve a fast and accurate quantitative understanding of the new virus structures to complement the sparse and inconsistent *in vitro* experiments with high-infectious virus cultivation.

Methods and Materials

Molecular Dynamics Parameters

Our MD experiments are conducted on the AiMOS supercomputer, a heterogeneous system of IBM POWER9 CPUs and Nvidia V100 GPUs. The S-protein data is obtained from the protein data bank (PDB: 6VXX). CHARMM27 is employed for describing the system of the S-protein and the SPC/E water molecules. The initial structure of the S-protein $12 \times 13 \times 16 \text{nm}^3$ was immersed in a cubic water box of $21 \times 21 \times 21 \text{nm}^3$. The periodic boundary condition is applied to all three boundaries of the water box. The total number of atoms is 805,218, of which 45,156 (5.6%) for the S-protein and 760,047 (94.4%) for water. Simulations are performed using the canonical (NVT) ensemble with a time step size of 2.5 femtoseconds. Each simulation starts with the first 6.5ns of

Fig. 1 Data processing pipeline from collection to analysis

energy minimization using gradient optimization, then followed a 0.2ns of MD runs to make the simulation reach a stable state for subsequent μ s-scale production runs.

Data Collection and Analysis

The raw trajectory obtained from a simulation was processed by aligning the protein in the center of the box in each frame. Then, each protein frame was aligned so that the protein was superimposed on the reference structure by overall protein rotation and translation. Next, the processed trajectory of atomic positions was used for the RMSD of backbone calculation. The number of mainchain-mainchain (M-M) and protein-water (P-W) H-bonds and SASA for the whole protein were calculated to measure the S-protein's thermal stabilities. Next, the unsupervised learning method, K-means was performed based on those measurements to cluster different temperatures into groups of viral survival durations. The deviation of the RBD was calculated to compare the spike opening between temperatures. RMSF was applied to measure the fluctuations of individual residues, and structural divergence for each residue was calculated. Fig. 1 outlines the flowchart of our data analysis.

Statistical Analysis: To obtain a comprehensive conclusion of the temperature impact, we performed a statistical analysis based on multi-perspective measurements. The *in silico* data are used as inputs of an unsupervised learning method based on K-means to cluster different temperatures into 4 groups of viral survival durations, second, minute, day, and week-long survival durations accordingly. In this, the *in silico* data include the average, standard deviation, kurtosis, and skewness of RMSD, as well as the average and variation of the P-W H-bonds.

Conformational Changes of RBD Analysis: The RBD (Residues Arg319-Phe541) of the S-protein is a potential target for vaccination and therapeutic development [23]. It is conserved among human coronaviruses and binds to the ACE2 receptor found in the lung tissues. The Residues Ser438-Gln506 of the RBD domain comprises the receptor-binding motif that interacts directly with ACE2 [23]. To measure the ACE2 binding competency of the S-protein at different temperatures, the opening is quantified in terms of how far the center of mass of an RBD deviates from its position in the first frame state (closed state).

$$D_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}$$

where D_i is the deviation of RBD at frame i , (x_i, y_i, z_i) are the coordinates of the center of mass of an RBD at frame i , and (x_0, y_0, z_0) the center of mass of an RBD at the first frame in the simulation.

Results and Discussions

We present our key observatories of the thermal properties of the S-protein in terms of RMSD, the number of H-bonds, and SASA. We use Fig. 2, a color scheme for temperature from 3°C to 95°C, to illustrate our results where the color-temperature association is obvious.

Fig. 2 Color representation for different temperatures

The average structure within the last 0.5 μ s at 37°C is taken as the reference structure (in gray) with which the structures at other temperatures are compared. As shown in Fig. 3, a first surprising observation is the size of S-protein shrinkage as temperature increases.

Fig. 3 Molecular superimposition of the average structure of different temperatures (top, colored) over the structure at 37°C (bottom, gray) where the color presents structural deviation in nm

The Thermal Properties of the S-Protein

The temperature-dependent variations of the S-protein structures are explored in terms of RMSD of the backbone (Fig. 4 R1-R3), the number of P-W and M-M H-bonds (Fig. 4 H1-H3), and SASA of the whole protein (Fig. 4 S1-S3).

The RMSD: The evolution of the protein structure was monitored along the MD trajectory by calculating RMSD taking the initial structure as the reference frame. During the MD simulations, the protein coordinates were recorded every 0.1ns and the RMSD of each frame was calculated from the

average RMSD of backbone atoms. Then, the moving average with 0.01 μ s window size was plotted against the simulated time in Fig. 4 R1. The distribution and box plot of the 1.5-3 μ s RMSD is shown in Fig. 4 R2 and R3 respectively.

The results in Fig.4 R1-R3 show:

- O.1** The S-protein shows greater stability (a smaller RMSD) at low temperatures (3°C) and is becoming more and more unstable (a larger and wider RMSD) as the temperature increases (significantly over 60°C). Quantitatively, its RMSD level starts at 3.7Å at 3°C, and gradually increases by ~35.1% to 5Å at 20°C-60°C, but it is dramatically doubled to 6~7Å at 70°C-95°C.
- O.2** It could be possible for the S-protein to stay long-term stable at cold temperatures by comparing the RMSD of 3.7Å at 3°C and 4.9Å at 20°C (the latter is 32.4% higher than the former). This *in silico* result is also in agreement with the reported case of salmon-attached coronavirus at 4°C that remains infectious for more than one week [2]. The study of the stability of SARS-CoV-2 also shows that the virus is stable at 4°C with only ~0.7-log reduction with infectious titers on Day-14 [14].
- O.3** The S-protein is becoming more sensitive to higher temperatures after 60°C. In particular, there appears a noticeable jump at RMSD from 5.3Å (60°C) to 6.2Å (70°C) that may serve as a signal for the presence of a critical heat denaturation around 60°C. To compare, the laboratory results [14] also showed a similar conclusion:

the viral inactivation time is greatly reduced to 5 mins with the incubation temperature increased to 70°C. Here, our *in silico* data showed a good agreement with this observation in this laboratory result.

The number of H-bonds: The P-W and M-M H-bond counts versus simulation time are presented in Fig. 4 H.1. Using 1.5-3 μ s data, the distributions for the P-W and M-M H-bond counts are presented in Fig. 4 H.2. To further explore the relationship between H-bond counts and the ambient temperature, the box plot of each type of H-bond is shown in Fig. 4 H.3. The results in Fig.4 H.1-H. 3 show:

- O.4** The average number of P-W H-bonds decreases quite linearly by 22.1%, from 6,136 to 5,026, as the temperature increases from 3°C to 95°C.
- O.5** The P-W H-bond counts at different temperatures present same groups as **O.1**: 3°C (6,136), 20°C (5,892) and 37°C (5,777), 60°C (5,486), 70°C (5,296) and 80°C (5,242), and 95°C (5,026). The count decreases from group to group are higher than 3% (4.0% from 3°C to 20°C, 5.0% from 37°C to 60°C, 3.5% from 60°C to 70°C, and 4.1% from 80°C to 95°C).
- O.6** The effect of increasing temperatures from 3°C to 80°C on the mainchain is insignificant. Quantitatively, compared with the decrease of P-W H-bonds (14.6% from 3°C to 80°C), the average number of M-M H-bonds only decreases by 3.6% from 3°C (1,257) to 80°C (1,212).

Fig. 4 The RMSD evolving as a function of time (R.1), distribution (R.2) and the box plot of the last 1.5-3 μ s values (R.3); the number of P-W (upper portion) and M-M (lower portion) H-bonds evolving with time (H.1) distribution (H.2) and the box plot of the last 1.5-3 μ s values (H.3); the SASAs evolving as a function of time (S.1), distribution (S.2) and the box plot of the last 1.5-3 μ s values (S.3)

The SASA: The SASA of the whole S-protein is calculated to explore the S-protein's conformational change. The SASA versus simulation time is shown in Fig. 4 S.1. Using 1.5-3 μ s data, the distributions and box plot for the SASA are presented in Fig. 4 S.2 and S.3, respectively. The results show:

O.7 The SASA at different temperatures present same groups as **O.1** and **O.5**: 3°C (1,281), 20°C (1,262) and 37°C (1,258), 60°C (1,221), 70°C (1,169) and 80°C (1,195), and 95°C (1,159). Among those groups, the average SASA at 37°C is 2.9% higher than it at 60°C and it at 60°C is 4.4% higher than it at 70°C.

Correlation between measurements: To correlate all *in silico* measurements applied to demonstrate the thermal properties, we intend to explore their sensitivities as shown in Fig. 5.

- The number of P-W H-bonds is positively correlated with the SASA (correlation coefficient 0.971), revealing a fact that to be accessible to the solvent is the precondition to form the H-bonds between water and protein. As the temperature increases, there are fewer P-W H-bonds, and the S-protein compactifies with lower SASA.
- The RMSD is negatively correlated with the number of M-M H-bonds (correlation coefficient -0.720); the more M-M

H-bonds there are, the more stable the S protein, *i.e.*, the lower the RMSD.

We achieved overwhelming consistency among measurements and, as they result from a single source, *i.e.*, the atom trajectories, it should not be a surprise.

Correlating *in silico* Measurements with Virus's Survival Times

The *in silico* data are used as inputs of an unsupervised learning method to cluster different temperatures into different groups of viral survival durations. In this, the *in silico* data include the average, standard deviation, kurtosis, and skewness of RMSD, as well as the average and variation of the P-W H-bond counts. A moving window statistic is used with a window size of 100ns and a stride size of 10ns. The K-means method is employed as the cluster with $n=4$. To compare with *in vitro* data, we correlate the four categories as second, minute, day, and week-long survival durations accordingly.

Based on *in vitro* data including the number of H-bonds and the RMSD, we can group the temperatures (**O.1**, **O.5** and **O.7**). Meanwhile, the survival times of the virus at different incubation temperatures were reported by *in vitro* experiments

Fig. 5 Correlation among measurements

[14]. The number of P-W H-bonds is investigated together with RMSD for the clustering of conformational behaviors at different temperatures. The samples are clustered by K-means in terms of the life duration of the virus based on 6 features captured from RMSD and the number of P-W H-bonds (constants for each temperature). The samples are grouped into four clusters which are colored by green, blue, orange and red, shown in Fig. 6.

Fig. 6 Clustering results of K-means using H-bond counts and average RMSD

Conformational Changes of RBD Analysis

In this section, the analysis and discussions are focused on the RBD of the S-protein. RBD is buried and inaccessible to receptors in a close state. Open state is necessary for binding with the receptors and to be infectious [31]. The opening is quantified in terms of how far the center of mass of an RBD deviates from its position in the close (or down) state [30]. A larger deviation of RBD usually implies a larger probability to bind. S-proteins that are more likely to adopt structures that are competent to bind ACE2 will be more likely to adopt highly open structures. Fig. 7 implies the following results:

O.8 The S-protein is more likely to adopt open structures at 37°C compared to 5 other temperatures with the deviation of the RBD up to 8Å. The RBD in chain A starts to be accessible to receptors rather than buried by NTD, indicating that the virus might be more competent to bind ACE2 at 37°C according to the *in silico* results.

O.9 At 3°C, the virus that can survive for over 2 weeks [14] is less infectious. The *in silico* result: at this temperature, the small RBD deviations of 1.5Å implied that RBD became hardly accessible to receptors for most of the time. For example, the virus stays active for over two weeks at 4°C [14] but could be less infectious at cold temperatures than normal ones.

Fig. 7 The distribution of RBD deviations in chains A, B and C (from top to bottom) at 3°C, 20°C, 37°C, 60°C and 70°C

Table 1 Clustering of survival times based on conformational information of the S-protein

Group	Temperatures	Features Descriptions
Group 1 (cold)	3~4°C	Blue group (Fig. 6): weeks of survival time. <i>In vitro</i> data: the virus at 4°C can survive for more than 2 weeks [14].
Group 2 (normal)	20°C to 37°C	Green group (Fig. 6): day-scale survival time. <i>In vitro</i> data: 7 days at 20°C and 1 day at 37°C [14].
Group 3 (high)	60°C to 70°C	Orange group (Fig. 6): minute-scale survival time. <i>In vitro</i> data: 10 mins at 56°C and around 1 min at 70°C [14].
Group 4 (very high)	Above 70°C to 95°C	Red group (Fig. 6): second-scale survival time. No <i>in vitro</i> data at this scale.

Residue-based Analysis

RMSF is used as a measurement of residue-based fluctuations at different temperatures. In Fig. 8, each point represented the RMSF of a specific residue ordered by its ResID in the sequence of the entire protein, and each column shows the result of a specific chain from A to C. As usual, a larger RMSF

infers a larger fluctuation of a residue and thus more active at a given temperature. Fig. 8 shows:

O.10 At 37°C, the chain A residues (an average RMSF at 1.64Å) are more flexible than the other two chains (1.40Å for chain B and 1.49Å for chain C) while at 80°C, most of the residues in chain B (2.05Å) and chain C (2.03Å) are more flexible than those in chain A (1.63Å).

Fig. 8 RMSF for each residue on three chains within the last 0.5μs simulation. The values of the top 25% are marked in orange, 25%-50% in light blue, 50%-75% in blue, and the rest in purple

- O.11** The NTD in chain A is more flexible than it in chain B and chain C at 20°C, 37°C, 80°C and 95°C. The NTD in chain A is most flexible (RMSF = 2.2Å) at 37°C among temperatures from 3°C to 95°C (around 1.7Å).
- O.12** the RBD in chain C is more flexible at 20°C and 37°C around (2.1Å) among temperatures from 3°C to 95°C (around 1.6Å).

Structural Divergence Analysis

The RMSF for each residue is measured in

$$RMSF_i = R_i = \sqrt{\frac{1}{T} \sum_{t=1}^T |r_i(t) - r_i(0)|^2}$$

A new entropy-like quantity is used to measure the structural divergence, at residue level, between two comparing temperatures T and T_0

$$SD_i(Temp, Temp_0) = R_i(Temp) \log \frac{R_i(Temp)}{R_i(Temp_0)}$$

where $R_i(Temp)$ is the RMSF at temperatures $Temp$ for residue i . Like in all definitions of entropy, the logarithm measures the relative (structural) divergence of a given residue at two different conditions (temperature, in our case). Simply, if $SD_i(Temp, Temp_0) > 0$, the residue i deviates from the reference structure at temperature $Temp$. This measure enables us to single out residues that are responsible for the overall structural change.

The reference structure can be fairly arbitrary to choose although, in most analyses, we set $Temp_0 = 37^\circ\text{C}$, for obvious reasons.

We calculate the structural divergence in the interesting temperature zone of 60°C-80°C. To spot the outliers of residues that cause significant structure divergence, we set $SD_i(Temp, Temp_0) = 0.07$ as the threshold with which we found the outliers for 60°C-80°C listed in Tables 2-4. The specific positions of the residue distribution in the protein are shown in Fig. 9. The dark yellow represents the S1 domain, and the light yellow represents the S2 domain. The cluster structure marked with balls are residue outliers, and its red, green, and blue colors represent chain A, B, and C respectively. The outliers' clusters for 80°C are zoomed shown in Fig. 10.

The Fig. 9-10 and Table 2-4 results shows:

- 1) At 60°C (11, 13, and 18 outliers in chain A, B, and C respectively) and 70°C (22, 15, and 34 outliers in chain A, B, and C respectively), outliers are evenly distributed in the entire S-protein.
- 2) At 80°C, among the 131 outliers in chain B, 45 are concentrated in RBD, 40 out of 45 forming a cluster shown in Fig. 10 (3). Moreover, 34 out of the 131 outliers in chain B are concentrated in NTD, and 33 out of 34 constituted a cluster shown in Fig. 10 (2). 51 out of 91 outliers form a cluster in chain C (Figure 10 (5)) and are concentrated on the carbon chain of S2.
- 3) We capture the large number of outliers in 80°C forming clusters caused 80°C RMSD higher than other temperatures.

Table 2 Chain A structural divergence outliers list

Chain A	S1 Domain Residue ID S2 Domain Residue ID
60	Arg214 Phe318 Arg408 Thr415 Lys529 Ser530 Pro812-Arg815 Leu858
70	Tyr451 Phe497 Gln498 Phe643-Ala647 Ser810-Lys811 Ser813-Lys814 Phe817 Phe855-Leu861 Thr912 Asn914
80	Cys136-Asn137 Asp405 Glu406 Arg408-Tyr423 Tyr505 Ser704 Thr747 Gly858

Table 3 Chain B structural divergence outliers list

Chain B	S1 Domain Residue ID S2 Domain Residue ID
60	Tyr369 Thr415 Ile418 Phe464 Arg466 Asn616 Glu619 Val620 Asn641 Val642 Gln644-Arg646
70	Ala27 Tyr28 Lys41 Phe135 Phe140 LEU212 Arg357 Arg466 Ile468 Tyr489 Arg815 Phe823 Lys825 Val826 Tyr904
80	Ala27-Asn30 Phe32 Thr33 Phe58-Ser60 Trp64 Phe65 Ala67 His69 Thr95 Glu96 Ser98 Ile100 Asn122-Thr124 Phe135 Asp138 Phe140-Val143 Asn165 Tyr170 Arg190 Phe220 Leu244 His245 Ala263 Thr315 Asn317-

	Arg319 Tyr351 Ser371 Ala372 Asp398 Phe400 Arg403 Arg408-Ile410 Ala411-Ala419 Tyr421 Tyr423 Phe429 Ala435 Arg454 Lys462 Arg466 Tyr489-Phe497 Asn501 Val503-Tyr505 Tyr508 Val512 Asp568-Asp571 Phe592 Thr791-Pro793 Pro807-Phe817 Asp820 Asn824 Lys825 Ile1081-Phe1089 Arg1091 Ile1114-Ile1130 Ile1132-Asn1135
--	---

Table 4 Chain C structural divergence outliers list

Chain C	S1 Domain Residue ID S2 Domain Residue ID
60	Phe133 Thr208-Ile210 Ala372 Ser375 Lys417-Ala419 Tyr449 Tyr451 Gly504 Thr645 Arg646 Val705 Lys795 Phe855 Arg995
70	His69 Asp80 Asn81 Lys97 Thr108 Leu110-Ser112 Thr114 Gln115 Phe133-Phe135 Cys166 Ile210-Val213 Asp215-Phe220 Ser371 Tyr495 Phe643 Thr645 Arg646 Lys795 Lys811 Tyr1110 Asp1118 Asn1135
80	Ala27 Thr29 Leu84 Pro85 Asn87 Lys97 Ser98 Ile100 Gly107 Thr108 Ile210 Asn211 Asp215-Phe220 Asn234 Thr240 Tyr269 Phe318 Arg319 Phe374 Ile468 Gly545 Phe592-Gly594 Val642-Arg646 Val705 Lys795 Gly799 Phe800 Phe802 Gln804 Pro807-Lys811 Lys814-Thr827 Lys854-Asn856 Thr912-Leu938 Gln965

Fig. 9 Identifications of residue outliers causing protein structure divergences and these outliers for experiments with 60-80°C to be zoomed further. (In this and the zoomed pictures, the colors are not related to the temperature color scheme.)

Fig. 10 The residue outlier clusters zoomed structure for 80°C, (1) RBD outlier cluster in chain A, (2) NTD outlier cluster in chain B, (3) RBD outliers cluster in chain B, (4) S2 outliers cluster in chain B, (5) S2 outliers cluster in chain C

Conclusions

Our simulations of μ s-long atomic resolutions demonstrated the S-protein's conformational changes affected by temperature varying and enabled us to infer the existence, and the possible critical temperature for the infectious of SARS-CoV-2 with evidence: **1)** The S-protein of SARS-CoV-2 is structurally stable at 3°C (**O.2**), agreeing with the *in vitro* results that the virus at 4°C can survive for more than 2 weeks [14], and is high-temperature resistance up to around 60°C (Fig. 4), under which most enzymes in the human body will be inactivated very soon; **2)** The *in silico* results indicate that all the measurements (RMSD, the number of P-W H-bonds and SASA) show a consistent signal that the heat effect difference between 60°C and 70°C is significant (**O.3**, **O.5** and **O.7**); **3)** By structural divergence analysis, we can see that in μ s-scale 80°C are broken first in the RBD, NTD, and carbon chain of the S2 domain.

Moreover, our *in silico* results could be correlated with the published *in vitro* results to correlate ambient temperature with the life duration and infectivity of the virus: **3)** The temperatures from 3°C to 95°C are clustered into 4 groups by unsupervised learning algorithm and the clustering result agrees with *in vitro* viral survival time scale from weeks to seconds [14] (Fig. 6); **4)** unlike the stability affected by increasing temperature, the opening of RBD is larger at 37°C while it is low with a low deviation of the RBD at cold temperature (3°C), room temperature (20°C) and high temperatures (higher than 60°C), indicating at 37°C the S-protein is more likely to adopt open structure although the life duration of the virus is shorter than it at 3°C.

For the first time, we present the μ s-scale MD studies of the temperature-varying conformation of a life-threatening S-protein. Our simulations on the most powerful IBM supercomputers, helping achieve fast and accurate quantitative understanding of the new virus structures, corroborate well with the published *in vitro* experiments, and, conversely, complement the sparse and inconsistent *in vitro* experiments with high-infectious virus cultivation.

One bottleneck of the current all-atom simulation is the computing power including computing resources and algorithms. To improve the simulation speed, a multiscale model concurrently considering components at its own characteristic scales [32] and an intelligent time stepping algorithm [33] can effectively relieve the computing load and shorten the simulation time, enabling long-term study with affordable efforts.

Acknowledgments The project is supported by the SUNY-IBM Consortium Award, PI: Y. Deng. All simulations were conducted on the AiMOS supercomputer at Rensselaer Polytechnic Institute and the WSC Cluster at the

IBM T. J. Watson Research Center through an IBM Faculty Award FP0002468 (PI: Y. Deng).

Funding The project is sponsored by Stony Brook University's OVPR & IEDM COVID-19 Seed Grant, PIs: P. Zhang, Y. Deng, M. Rafailovich, and M. Simon.

Availability of data and material All original data are available upon request.

Code availability Data generation is done through the open-source Gromacs while data analysis is done by our own code that is available upon request.

Author Contributions Meichen Song: Setting up and performing simulations, Data collection and analysis, Conceptualization, Writing- Original draft. Ziyuan Niu: Data analysis, Conceptualization, Writing- Original draft. Karin: Discussion. Yuefan Deng: Resources, Supervision, Writing- Editing. Marcia Simon: Resources. Miriam Rafailovich: Resources. Peng Zhang: Data analysis, Resources, Supervision, Writing- Reviewing and Editing.

Declarations

Ethics approval The manuscript is prepared in compliance with the Ethics in Publishing Policy as described in the Guide for Authors.

Consent to participate The manuscript is approved by all authors for publication.

Consent for publication The consent for publication was obtained from all participants.

Conflict of interest The authors declare that they have no conflict of interest

References

1. Sarkar K, Khajanchi S, Nieto JJ (2020). Modeling and forecasting the covid-19 pandemic in india. *Chaos Solitons Fractals* 139: 110049. 10.1016/j.chaos.2020.110049
2. Dai M, Li H, Yan N, Huang J, et al. (2020). Long-term survival of salmon-attached sars-cov-2 at 4° c as a potential source of transmission in seafood markets. *bioRxiv* 10.1101/2020.09.06.284695
3. Rizou M, Galanakis IM, Aldawoud TM, Galanakis CM (2020). Safety of foods, food supply chain and environment within the covid-19 pandemic. *Trends in Food Science & Technology* 102: 293-299.
4. Rubin D, Huang J, Fisher BT, Gasparrini A, et al. (2020). Association of social distancing, population density, and temperature with the instantaneous reproduction number of sars-cov-2 in counties across the united states. *JAMA Netw Open* 3(7): e2016099. 10.1001/jamanetworkopen.2020.16099
5. Sobral MFF, Duarte GB, da Penha Sobral AIG, Marinho MLM, et al. (2020). Association between climate variables and global transmission of sars-cov-2. *Science of The Total Environment* 729: 138997.
6. Tosepu R, Gunawan J, Effendy DS, Lestari H, et al. (2020). Correlation between weather and covid-19 pandemic in jakarta, indonesia. *Science of the total environment* 725: 138436.
7. Darnell ME, Subbarao K, Feinstone SM, Taylor DR (2004). Inactivation of the coronavirus that induces severe acute respiratory syndrome, sars-cov. *Journal of virological methods* 121(1): 85-91.
8. Leclercq I, Batejat C, Burguière AM, Manuguerra JC (2014). Heat inactivation of the middle east respiratory syndrome coronavirus. *Influenza and other respiratory viruses* 8(5): 585-586.
9. Millet JK, Kien F, Cheung CY, Siu YL, et al. (2012). Ezrin interacts with the sars coronavirus spike protein and restrains infection at the entry stage. *PLoS One* 7(11): e49566. 10.1371/journal.pone.0049566

10. Song W, Gui M, Wang X, Xiang Y (2018). Cryo-em structure of the sars coronavirus spike glycoprotein in complex with its host cell receptor ace2. *PLoS Pathog* 14(8): e1007236. 10.1371/journal.ppat.1007236
11. Sturman LS, Ricard CS, Holmes KV (1990). Conformational change of the coronavirus peplomer glycoprotein at ph 8.0 and 37 degrees c correlates with virus aggregation and virus-induced cell fusion. *J Virol* 64(6): 3042-3050. 10.1128/JVI.64.6.3042-3050.1990
12. Biryukov J, Boydston JA, Dunning RA, Yeager JJ, et al. (2020). Increasing temperature and relative humidity accelerates inactivation of sars-cov-2 on surfaces. *MSphere* 5(4): e00441-00420.
13. Chan KH, Sridhar S, Zhang RR, Chu H, et al. (2020). Factors affecting stability and infectivity of sars-cov-2. *J Hosp Infect* 106(2): 226-231. 10.1016/j.jhin.2020.07.009
14. Chin AW, Chu JT, Perera MR, Hui KP, et al. (2020). Stability of sars-cov-2 in different environmental conditions. *The Lancet Microbe* 1(1): e10.
15. Matson MJ, Yinda CK, Seifert SN, Bushmaker T, et al. (2020). Effect of environmental conditions on sars-cov-2 stability in human nasal mucus and sputum. *Emerg Infect Dis* 26(9): 2276. 10.3201/eid2609.202267
16. Riddell S, Goldie S, Hill A, Eagles D, et al. (2020). The effect of temperature on persistence of sars-cov-2 on common surfaces. *Virol J* 17(1): 145. 10.1186/s12985-020-01418-7
17. Wang T, Lien C, Liu S, Selveraj P (2020). Effective heat inactivation of sars-cov-2. *medRxiv*: 2020.2004.2029.20085498. 10.1101/2020.04.29.20085498
18. Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, et al. (2020). Aerosol and surface stability of sars-cov-2 as compared with sars-cov-1. *New England Journal of Medicine* 382(16): 1564-1567.
19. Li F (2016). Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 3(1): 237-261. 10.1146/annurev-virology-110615-042301
20. 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(3): 254. 10.3390/v12030254
21. de Groot RJ, Luytjes W, Horzinek MC, van der Zeijst BA, et al. (1987). Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J Mol Biol* 196(4): 963-966. 10.1016/0022-2836(87)90422-0
22. Walls AC, Park YJ, Tortorici MA, Wall A, et al. (2020). Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell* 181(2): 281-292 e286. 10.1016/j.cell.2020.02.058
23. Lan J, Ge J, Yu J, Shan S, et al. (2020). Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor. *Nature* 581(7807): 215-220. 10.1038/s41586-020-2180-5
24. Tian X, Li C, Huang A, Xia S, et al. (2020). Potent binding of 2019 novel coronavirus spike protein by a sars coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect* 9(1): 382-385. 10.1080/22221751.2020.1729069
25. Verkhivker GM (2020). Molecular simulations and network modeling reveal an allosteric signaling in the sars-cov-2 spike proteins. *J Proteome Res* 19(11): 4587-4608. 10.1021/acs.jproteome.0c00654
26. Komatsu TS, Okimoto N, Koyama YM, Hirano Y, et al. (2020). Drug binding dynamics of the dimeric sars-cov-2 main protease, determined by molecular dynamics simulation. *Scientific reports* 10(1): 1-11.
27. He J, Tao H, Yan Y, Huang SY, et al. (2020). Molecular mechanism of evolution and human infection with sars-cov-2. *Viruses* 12(4): 10.3390/v12040428
28. Zhou Z, Yang Z, Ou J, Zhang H, et al. (2021). Temperature dependence of the sars-cov-2 affinity to human ace2 determines covid-19 progression and clinical outcome. *Comput Struct Biotechnol J* 19: 161-167. 10.1016/j.csbj.2020.12.005
29. Rath SL, Kumar K (2020). Investigation of the effect of temperature on the structure of sars-cov-2 spike protein by molecular dynamics simulations. *Front Mol Biosci* 7: 583523. 10.3389/fmolb.2020.583523
30. Zimmerman MI, Bowman G (2021). Sars-cov-2 simulations go exascale to capture spike opening and reveal cryptic pockets across the proteome. *Biophysical Journal* 120(3): 299a.
31. Vankadari N, Wilce JA (2020). Emerging wuhan (covid-19) coronavirus: Glycan shield and structure prediction of spike glycoprotein and its interaction with human cd26. *Emerg Microbes Infect* 9(1): 601-604. 10.1080/22221751.2020.1739565
32. Zhu Y, Zhang P, Han C, Cong G, et al. (2021). Enabling ai-accelerated multiscale modeling of thrombogenesis at millisecond and molecular resolutions on supercomputers. *International Conference on High Performance Computing*, Springer. 237-254.
33. Han C, Zhang P, Bluestein D, Cong G, et al. (2021). Artificial intelligence for accelerating time integrations in multiscale modeling. *Journal of Computational Physics* 427: 110053.