Spike conformation transition in SARS-CoV-2 infection

Liaofu Luo* (College of Physical Science, Inner Mongolia University, Hohhot, China)

Yongchun Zuo (College of Life Sciences, Inner Mongolia University, Hohhot, China)

* Correspondence address: lolfcn @ imu.edu.cn

Abstract A theory on the conformation transition for SARS-CoV-2 spike protein (S) is established. The conformation equilibrium between open (up) and closed (down) conformations of receptor binding domain (RBD) is studied from the first-principle. The free energy change in conformation transition of S protein is introduced and we demonstrated that it includes two parts, one from the difference of conformation potential and another from the variation of structural elasticity. The latter is dependent of amino acid mutation. When the amino acid mutation of S protein causes a substantial reduction of elastic energy the equilibrium is biased to the open conformation. Only then can the virus infection process continue. That both the D614G mutation and the K986P mutation increase the COVID-19 infectivity and why a large number of mutations, including those at interface residues, have not been selected in current SARS-CoV-2 pandemic are interpreted from the presented theory. The evolution of coronavirus dependent on the alteration of conformation equilibrium is indicated. Finally, introduction of electric field to change the conformation potential barrier and how the conformation equilibrium depends on temperature and humidity are briefly discussed.

1 Introduction

The current SARS-CoV-2 pandemic has created urgent needs for diagnostics, therapeutics and vaccines. However, meeting these needs firstly requires a deep understanding of the mechanism of viral infection. For all enveloped viruses, membrane fusion is a key early step for entering host cells and establishing infection. The surface of coronaviruses are decorated with a spike (S) protein (about 1273 amino acids for SARS-CoV-2), a large class 1 fusion protein. The S protein forms a trimeric complex that can be functionally categorized into two distinct subunits S1 and S2 subunits. There is a receptor binding domain (RBD) in S1 (between sequence sites 331 to 528 for SARS-CoV-2) which interacts with a
host-cell receptor protein to trigger membrane fusion. The host-cell receptors for SARS-CoV and MERS-CoV are angiotensin-converting enzyme 2 (ACE2) and DPP4 respectively [1][2]. Recent advances in cryo-electron microscopy (cryo-EM) characterization of the spike protein revealed that the RBDs adopted at least two distinct conformations. RBD can be either in the open or in the closed position (called up or down conformation respectively).[3] In the up conformation, the RBD jut out away from the rest of S, such that they can easily bind with ACE2. In the down conformation the RBDs are tightly packed, preventing binding by ACE2.[4] The receptor-binding event may trap the RBD in the less-stable up conformation, leading to destabilization of S1, triggering the conformational change of S2 from prefusion to postfusion state and initiating the membrane fusion. The SARS-CoV cell entry also depends on transmembrane protease serine 2 (TMPRSS2) etc which help to cut S to units S1 and S2.[5]

The above conformational transition processes can be expressed through a set of equations

\[
\text{RBD(closed)} \leftrightarrow \text{RBD(open)} \quad (1)
\]

\[
\text{RBD(open)} + \text{ACE2} \rightarrow \text{AS} \quad (2)
\]

\[
S + E \rightarrow S_1 + S_2 + E
\]

\[
S_{2\text{ prefus}} \rightarrow S_{2\text{ postfus}} \quad (3)
\]

where AS in Eq (2) denotes the RBD-receptor bound state. Eq (3) is a set of equations causing the transition of S protein from prefusion to postfusion state and in its first equation E denotes the enzyme TMPRSS2. The di-directional transition of Eq (1) makes RBD attaining at an equilibrium between open and closed conformations. Eq (1) is the starting point of the viral infection, we shall focus on it in the paper.

2 Method - Mathematical relation of conformation equilibrium

Eq (1) describes the conformation transition of RBD. The closed/open transition of RBD proceeds in two directions the equilibrium of which determines the first step of viral infection. We shall study the conformation equilibrium between RBD(closed) and RBD(open). The conformational potential is expressed through Fig 1. Two minima separated by a potential barrier represent two conformations of RBD. In principle the RBD can jump from the left to the right, or vice versa from the right to the left and finally attains an equilibrium with a definite probability distribution of two conformations. Suppose A denotes closed conformation and B denotes open conformation. Generally, if the left minimum A is lower than the right B, then S1 takes inactive conformation A with large probability. As the spike protein located in this structure the subsequent steps of Eqs (2) and (3) will not start and the virus infectious process cannot continue. Oppositely, in
order to start Eq(2), the equilibrium of Eq(1) should be biased to the right open conformation.

Define the free energy increase $\Delta G$ (G in conformation B minus G in conformation A) one has [6]

$$\Delta G = -k_B T \ln \frac{Z_B}{Z_A}$$

(4)

Here $Z_A$ and $Z_B$ are partition function of the statistical system in state A and B respectively that is derived from the summation of Boltzmann factor over vibration states and $Z_A / Z_B$ means the ratio of the probabilities in two conformations A and B. We have [7]

$$\frac{Z_A}{Z_B} = e^{-\beta (V_A - V_B)} \quad (\beta = \frac{1}{k_B T})$$

(5)

$$Y_{A/B} = \frac{1}{\sqrt{k_B T}} \frac{1}{e^{\beta h \omega_B} - 1} \frac{1}{e^{\beta h \omega_A} - 1}$$

(6)

$Y_{A/B}$ comes from the ratio of the summation over vibration states around two minima. If the vibration is neglected then the probability ratio is simply determined by the symbol of $V_A$-$V_B$. Suppose $V_A < V_B$, A is the advantageous conformation due to $Z_A / Z_B > 1$. However, the vibrations around the minimum of potential are important for the localization of a definite conformation. From Eq (6) one has

$$Y_{A/B} = \frac{1}{\sqrt{k_B T}} \frac{1}{e^{\beta h \omega_B} - 1} \frac{1}{e^{\beta h \omega_A} - 1}$$

(6.1)

$$Y_{A/B} = \exp\left\{ \frac{h(\omega_B - \omega_A)}{2k_B T} \right\} \quad \text{as} \quad k_B T \ll h \omega_A, h \omega_B$$

(6.2)

In general

$Y_{A/B} > 1$ and decreasing monotonously with $T$ as $\omega_A < \omega_B$

$0 < Y_{A/B} < 1$ and increasing monotonously with $T$ as $\omega_A > \omega_B$

(7)

Inserting (6) into (5) we have

$$\Delta G = V_B - V_A + k_B T \ln Y_{A/B}$$

$$= V_B - V_A + \frac{h \omega_B}{2} - \frac{h \omega_A}{2} \quad \text{as} \quad k_B T \ll h \omega_A, h \omega_B$$

$$= V_B - V_A + k_B T \ln \frac{\omega_B}{\omega_A} \quad \text{as} \quad k_B T \gg h \omega_A, h \omega_B$$

(8)

The right-hand-side of Eq(8) contains two parts: $V_{A,B}$ describes the conformational energy, and $\omega$-related term $k_B T \ln Y_{A/B}$ describes the elastic energy. The meaning of $k_B T \ln Y_{A/B}$ can easily be seen from its expression in low-temperature limit since $h \omega_{A,B}/2$ is the zero-point energy of elastic vibration. Therefore, by taking the elastic
energy into account the conformation with lower vibration frequency increases its probability. When $\omega_B$ is much smaller than $\omega_A$, in spite of $V_A < V_B$ one always has $Z_A / Z_B < 1$ and the conformation B is the advantageous (with large probability) instead of A.

Another relation can be used in estimating the elastic energy term $Y_{A/B}$ is

$$k_B T \ln Y_{A/B} = \frac{h(\omega_B - \omega_A) - \text{ctnh} \frac{\omega_A}{2k_B T}}{\frac{\omega_A - \omega_B}{\omega_A}} \text{ (as } \left| \frac{\omega_A - \omega_B}{\omega_A} \right| \ll 1)$$

where the function ctnh is defined by $\text{ctnh} x = \frac{e^x + e^{-x}}{e^x - e^{-x}}$, an odd function decreasing with $x$ and always larger than 1 for positive $x$.

3 Results

The structural elasticity of spike protein is dependent of its amino acid sequence. One may increase or decrease the S elasticity energy through amino acid mutation. Therefore, the originally inactive conformation up can transform into an active advantageous one by selecting appropriate amino acid mutation. Recently the deep mutational scanning of SARS-CoV-2 RBD was completed and its constraints on folding and ACE2 binding were analyzed [8]. However, how the mutation influences the structural elasticity and conformation equilibrium and in turn, influences the virus infection, has not been studied. In the following we will sketch the main results of our method for explaining experimental facts.

3.1) D614G mutation

There is strong evidence that the spike protein mutation D614G increases infectivity of the COVID-19 virus and the G614 variant has become the most prevalent form in the global pandemic [9]. How to explain this important event in the virus evolution? The recent analysis on the structure and function of the D614G spike protein indicated that although the D614G affinity for ACE2 is reduced the conformation is shifted towards an ACE2 binding-competent state[10]. According to above-mentioned conformation equilibrium theory, due to the mutation D614G removing the H-bond between residues D614 in S1 and T859 in S2, the RBD structural elasticity has been changed and the flexibility has been increased. For the wild type spike the RBD is in closed conformation A (in left well of Fig 1). where $V_A$ is lower than $V_B$. However, the frequency $\omega_B$ (in right well of Fig 1) decreases largely due to amino acid replacement and the
potential minimum B changes to the advantageous conformation. It means the amino acid mutation triggers the transition of RBD from a closed to a more open conformation. The open RBD will be able to initiate the subsequent processes following Eq (2) and Eq(3) and consequently the spike protein changes itself from prefusion to postfusion states to mediate fusion of viral and cellular membrane. This interprets why the mutation D614G largely increases the infectivity of the COVID-19 virus.

3.2) K986 P mutation
In recent mRNA vaccine design people demonstrated that mRNA expressing SARS-CoV2S-2P is a potent immunogen. They identified 2 proline substitutions (2P) at the apex of the central helix and heptal repeat 1 (HR1) that effectively stabilized MERS-CoV and SARS-CoV proteins in the prefusion conformation [11]. The prefusion-stabilized protein immunogens that preserve neutralization-sensitive epitopes are an effective vaccine strategy for enveloped viruses. Here the key step is the 2P mutation K986 P and V987P in SARS-CoV2 spike sequence. The K986 P mutation removes a salt bridge between Lys986 and either ASP427 or ASP428 of another protomer in the trimer interface [2]. The mutation relaxes the structure of S protein. That is, due to the mutation the conformational potential curve in Fig 1 changes to more flat around its right minimum. Thus, S protein jumps from the left closed conformation A to the right open conformation B. Since B is an easily infectious conformation, the K986 P mutation provides an immunogen design that can trigger immediate rapid manufacturing of an mRNA vaccine.

3.3) The most commonly studied amino acid variants in spike
The global frequency of amino acid variants in sites of interest is given in Table 1. The corresponding free energy increase is also calculated and listed in the table.

| Mutation number | Spike mutation | Region            | Count | -ΔG/k_BT |
|-----------------|----------------|-------------------|-------|----------|
| 1               | D614G          | S1 CTD domain     | 71%   | 0.895    |
| 2               | L5F            | Signal peptide    | 0.6%  | -5.11    |
| 3               | R21I/K/T       | S1 NTD domain     | 0.5%  | -5.29    |
| 4               | A829T/S        | Fusion peptide    | 0.3%  | -5.81    |
| 5               | D839Y/N/E      | Fusion peptide    | 0.5%  | -5.29    |
| 6               | D936Y/H        | HR1               | 0.9%  | -4.70    |
| 7               | P1263L         | Cytoplasmic tail  | 0.7%  | -4.95    |
| 8               | K986P          | HR1               | >70%  | >0.85    |
| 9               | A475W/P        | RBD               | Small(<0.5%) | <5.3 |
| 10              | G476I/L/P      | RBD               | Small(<0.5%) | <5.3 |
Experimental data of mutations 1-7 are taken from [9]. Data of mutation 8 are taken from [2][11]. Data of mutations 9 and 10 are taken from [8]. $-\Delta G/k_BT$ in the last column is calculated by using Eq(8) where $Z_A/Z_B$ is taken from the experimental count. The lower bound of count in mutations 8 is estimated from thermostable spike trimer where about 80% of S-R/PP trimmers have one open RBD [12] and the upper bound of count in mutations 9 and 10 is assumed tentatively in the calculation.

From the table we find the free energy increase $\Delta G$ of open conformation relative to closed is positive in nearly all mutations apart from D614G and K986P. It explains why a substantial number of mutations, including those at ACE2 interface residues, have not been selected in current SARS-CoV-2 pandemic isolates. Moreover, Table 1 indicates $-\Delta G/k_BT$ always takes a positive value near +1 or takes a negative value near -5 for all mutations. The former means the amino acid mutation makes the elastic energy of open conformation much lower than the closed conformation so that $\Delta G$ changes to negative. The latter means the amino acid mutation having not brought any obvious change of elastic energy so that $\Delta G \cong (V_B - V_A)$ remains positive.

How to explain the relation between amino acid mutation and the change of spike elasticity? Set amino acids D,E and Y charges negative; H,K and R charges positive; and others are electric neutrality. One may assume a large change of elastic energy can be found only if the residue’s electric charge difference is 1 unit or more in the mutation and otherwise, if the average electric charge difference is obviously smaller than 1 unit then the elastic energy variation is too small so that the mutation cannot trigger the structural changes of S protein. From the experimental data in Tab 1 we know that in the former case $-\Delta G/k_BT$ takes a positive value near +1 and in the latter case $-\Delta G/k_BT$ takes a negative value near -5. Therefore the electricity change in amino acid mutation determines the elasticity change of the spike and finally determine which one (open or closed) is the advantageous conformation of RBD.

3.4) More conformations of virus S protein

Three conformations of the prefusion trimer observed by using cryo-electron microscopy on intact virions: all RBDs in the closed position; one RBD in the open position; and two RBDs in the open position.[3] However, the two-open conformation has only been observed in vitro after inserting multiple stabilizing mutations. Moreover, through the designed mutations in S protein these authors observed distinct closed and locked conformations of the S trimer and the classification of the cryo-EM data showed that the disulfide bond formation is beneficial to the closed RBDs.[12] The above experimental data can be explained as well by our model. Disulfide bond is a strong bond. The disulfide bond formation gives another factor to tighten the S protein structure in addition to commonly the H-bond and salt bridge. Following the present model the multiple stabilizing mutations make the conformational potential $U(\theta)$ of two minima generalized to a potential with three or more minima. The generalized multi-minima can describe the
abundant structures of S protein and many new conformations of RBD appeared in the experiment. Moreover, any change of the vibration frequency $\omega$ around one minimum will give additional contribution to the free energy of the prefusion trimer. The conformation equilibrium among multi-minima will provide a starting point for the study of the phase-transition dynamics of the system.

4 Discussions

4.1 Conformational equilibrium of spike protein in amino acid mutation. The structure and conformational change of the spike protein is of special significance for virus infection and transmission. To understand the virus entry and viral infection one needs not only the RBD expression and its ACE2-binding affinity, but also the conformational equilibrium of the spike protein in amino acid mutation. We found that the SARS-CoV-2 infection of humans and its pandemic is always achieved by amino acid mutation on the spike, commonly the mutation of acid residues or basic residues that triggers the loss of some hydrogen bonds or salt bonds and changes the elastic energy of the spike protein. The change of elastic energy in two conformations $A$ and $B$ is associated with the variation of conformational vibration frequency $\omega_A - \omega_B$ or $\omega_A / \omega_B$. The elastic energy is an important part of the free energy. Quantitatively speaking, as the amino acid mutation does not cause the breaking of hydrogen bond or salt bond, the free energy difference $\Delta G$ between closed and open conformations is about $-5k_B T$, while the amino acid mutation causing break of hydrogen bond or salt bond, $\Delta G$ is near $+1 k_B T$. The difference $6k_B T$ in two cases comes from the elastic energy variation. In other words, the hydrogen bond (salt bond) break in spike protein generally needs energy input about in the order of $6k_B T \sim 0.15$ ev which is in the reasonable range of the magnitude.

4.2 Evolution of SARS-CoV dependent on the alteration of conformation equilibrium due to amino acid mutation. SARS-CoV-2 RBD binds ACE2 with higher affinity than SARS-CoV-1. However, deep mutational scanning of all amino acid mutations gives new understanding on the problem, suggesting that there is a substantial mutational space consistent with sufficient affinity to maintain human infectivity [8]. Following our point of view, when we study how the amino acid mutation influences the virus evolution we should consider the conformation equilibrium apart from ACE2 affinity. Comparing RBD sequences between SARS-CoV-1 and SARS-CoV-2 we found there are 49 single amino acid mutations in total. Fourteen of these mutations have the electric charge variation of 1 unit. Eight mutations from SARS-CoV-1 to SARS-CoV-2 namely E354N, R439N, K452L, Y455L, D476G, K478T, D494S and Y498Q are charge-decreasing (from charge negative or positive to neutral), and six mutations V417K, N460K, V471E, F473Y, P484E and N519H are electric charge-increasing (from charge neutrality to negative or positive). These single mutations occurred in different periods of virus evolution. Assuming some of them can trigger the structure variation of the spike protein, the charge-decreasing mutation may loosen the structure and decrease the vibration frequency, while the charge-increasing mutation may tighten the structure and increase the vibration frequency. Therefore, the amino acid mutation caused changes
more than once on the structural elasticity of spike protein and the conformation equilibrium of RBD in the virus evolutionary history. It is interesting to note that two charge-decreasing mutations D476G and K478T are located exactly in the high entropy mutational cluster (from site 475 to 483) for SARS-CoV-2 [9]. One may assume that they occurred in the 2019 winter and are closely related to the 2019 Novel Coronavirus outbreak. If we further assume some electric charge-increasing mutations had happened in the past and triggered the structural variation then why SARS CoV-1 suddenly and quietly left from human in 2003 and after 17 years its variant SARS CoV-2 rapidly spread in the world might be explained.

4.3) Introduction of electric field to change the conformational barrier of the spike. As shown in Fig 1 there is a potential barrier between two conformations. The transmission coefficient is dependent of the height and width of the potential barrier. Introducing electric field will change the conformational barrier of the spike effectively and in turn change the distribution of RBD in two conformations. It was reported that as the electric field increases beyond 0.02 au, the net electron density starts to move from C-H bond towards the carbon, causing the bond to begin to weaken and lengthen. Thus, the static electric field of appropriate strength and direction can break some H-bond and salt bond in the spike and changes the conformation equilibrium of RBD.

4.4) Equilibrium between RBD closed and open conformations correlated with temperature and humidity. Conformational equilibrium of RBD is dependent of temperature, that can be seen from Eqs (4) and (5). From Eq (5) we know that $Y_{A/B}$ depends on temperature monotonously, decreasing with $T$ as $\omega_A<\omega_B$ and increasing with $T$ as $\omega_A>\omega_B$ (Eq(7)). Since $Y_{A/B}$ is related to the equilibrium constant $K_{eq}$ the temperature dependence of the virus entrance can be tested against experiments by measuring the equilibrium constant for the virus assembly.[13] Apart from temperature, the conformational equilibrium depends on humidity. It was reported that the soluble S trimer with the PP mutation has a looser structure than the full-length S with wild-type sequence [2]. So, the moisture may be conductive to K986 P mutation and makes the virus infectious. In fact, the virus can be modeled as a charged sphere. From the electrostatics for salty solution, one can arrive at an expression for the potential $U(\theta)$ at the surface of the charged sphere and the dielectric constant $\epsilon$ has entered into the expression of $U(\theta)$ [13]. It means the elastic frequency $\omega^2$ should be replaced by $\omega^2/\epsilon$. For water $\epsilon=80$. Therefore, the frequency parameter takes a reduced value $\omega/9$ in the full salty solution instead of $\omega$ in vacuum and it should be reduced by a multiple up to 9 in the humid environment. This gives a quantitative estimate on how the virus infection depends on humidity.

References
1. Wrapp D et al. Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies. Cell 181, 1004-1016, 2020.
2. Cai YF et al. Distinct conformational of SARS-CoV-2 spike protein. Science
3. Ke ZL et al, Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature 2020 https://doi.org/10.1038/s41586-020-2665-2.

4. Gui M et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. 27,119–129 (2017)

5. Hoffman M et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271-280, 2020

6. Landau LD, Lifshitz EM. Statistical Physics. Pergamon Press (1958).

7. Luo LF. Conformation dynamics of macromolecules. Int J Quantum Chemistry 32, 435-450 (1987).

8. Starr TN et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell, 11 August 2020. https://doi.org/10.1016/j.cell.2020.08.012

9. Korber B et al. Tracking changes in SARS-CoV-2 spike: Evidence that D614 G increases infectivity of the COVID-19 virus. Cell 182, 1-16, 2020

10. Yurkovetskiy, L., Wang, X., Pascal, K.E.et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant, Cell (2020), doi: https://doi.org/10.1016/j.cell.2020.09.032

11. Corbett KS et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature, 5 August 2020.

12. Xiong XL et al. A thermostable closed SARS-CoV2 spike protein trimer. Nature Struct Mol Biol.2020 https://doi.org/10.1038/s41594-020-0478-5.

13. Phillips R, Kondev J and Theriot J. Physical Biology of the Cell. GS Garland Science (2008)
Fig 1  Conformational Potential $U(\theta)$ versus $\theta$

$$U(\theta) = V_A + \frac{1}{2} I \omega_A^2 (\theta - \theta_A)^2 \quad \text{(left)}, \quad U(\theta) = V_B + \frac{1}{2} I \omega_B^2 (\theta - \theta_B)^2 \quad \text{(right)}$$

Two minima separated by a potential barrier represent two conformations A (left) and B (right) of RBD, $\theta$ is conformational coordinate, $\omega$ – the frequency parameter and $I$ – the inertia parameter describing the vibration around the minimum.