Electronic Supplementary Information

Development of Multiscale Ultra-Coarse-Grained Models for the SARS-CoV-2 Virion from Cryo-Electron Microscopy Data

Fengyu Li[1], Yuwei Zhang[1], Fei Xia[2*] & Xin Xu[1,3*]

[1]*Collaborative Innovation Center of Chemistry for Energy Materials, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, MOE Key Laboratory of Computational Physical Sciences, Departments of Chemistry, Fudan University, Shanghai 200433, China*

[2]*School of Chemistry and Molecular Engineering, NYU-ECNU Center for Computational Chemistry at NYU Shanghai, East China Normal University, Shanghai 200062, China*

[3]*Hefei National Laboratory, Hefei 230088, China*

Fig. S1 The visualization of the contacting strengths of the UCG beads as parameter *R* adopts (a) 10 Å, (b) 20 Å, and (c) 30 Å, respectively. The bond strengths from the strong to the weak are colored in the right bar. The total connection strength of the *i*th bead is defined as $G_i = \sum_{j \in \mathrm{UCG}} k_{ij}$. The value of R = 10 Å generally leads to the weak interactions as well as the value of 30 Å yields the too rigid bond connections. Thus, parameter *R* adopts the empirical value of 20 Å finally to give a reasonable description of the heterogeneity of the contracting strengths.

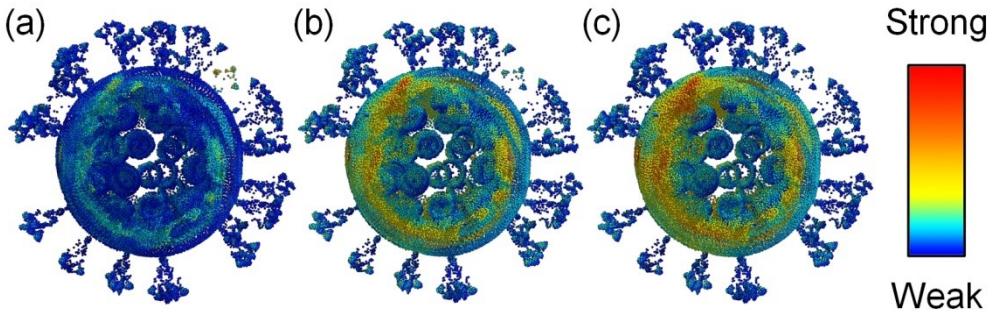

Fig. S2 Interactions at the interface of AA and UCG moieties. The Cα atoms in the tail of the AA spike protein and the UCG beads of the envelope are connected with the harmonic springs of 1.0 kcal·mol^{-1}·Å^{-2}. The harmonic bonds are virtualized as yellow sticks.

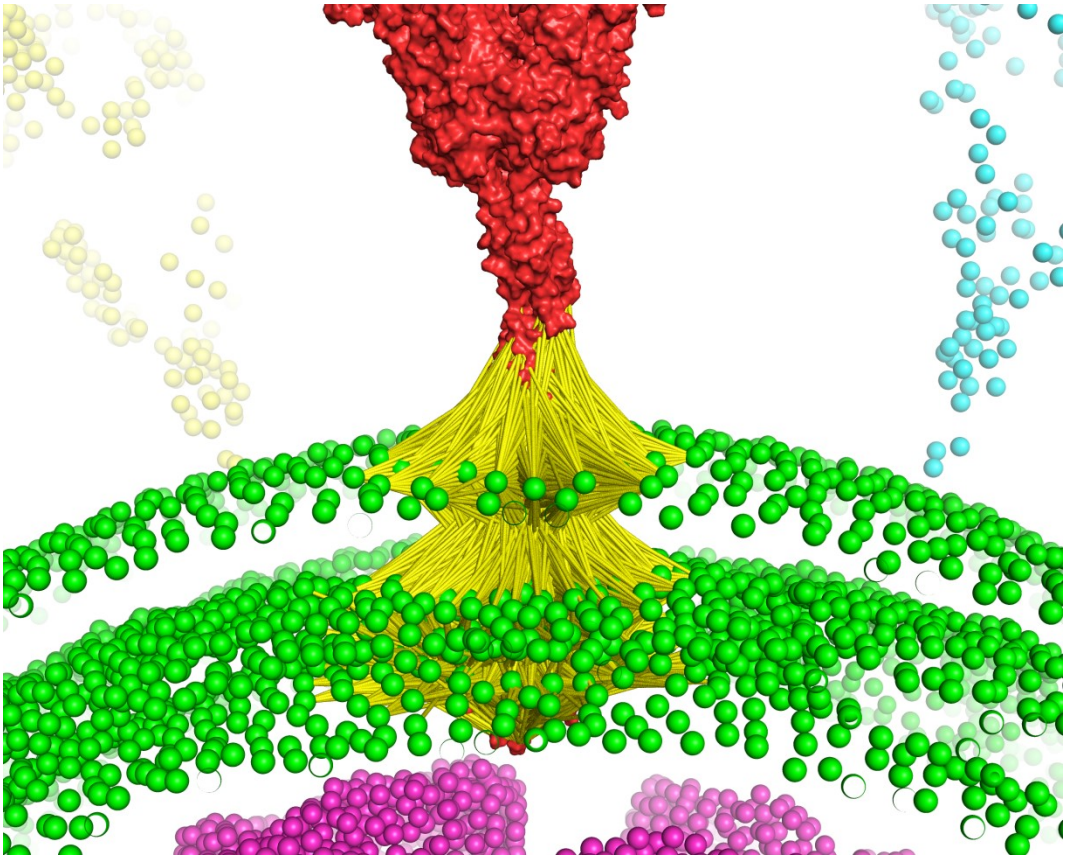

Fig. S3 Illustration of grouping technique where 300 two-dimension points (in cyan) are grouped into 114 discrete points (in red). Each raw value is adjusted to the nearby discrete value. The spans of the modified data along the two dimensions are both 0.1 in this example. Likewise, the pairwise force field parameters (k, r_0) resemble the two-dimension points, and can be reduced to smaller values to reduce memory costs.

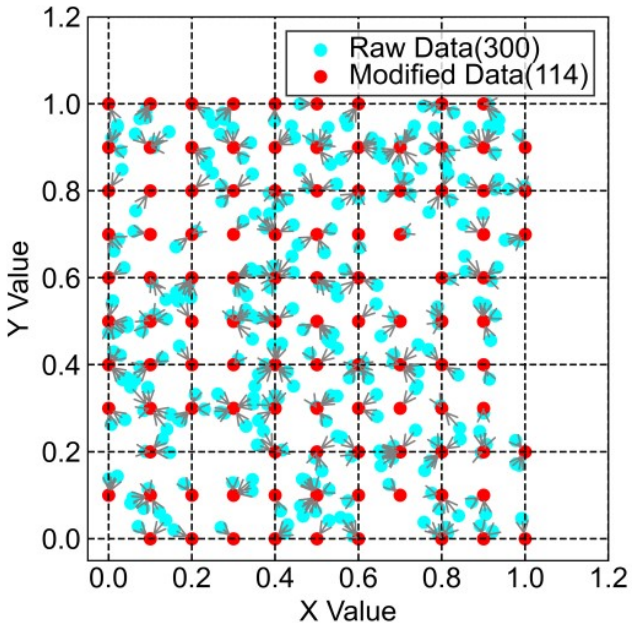

Fig. S4 Density distribution of the SARS-CoV-2 virion in the cryo-ET data (EMD-30430). The three contour iso-surfaces of virions are virtualized at the density values of 0.5, 15.0, and 100.0 (a.u.), respectively.

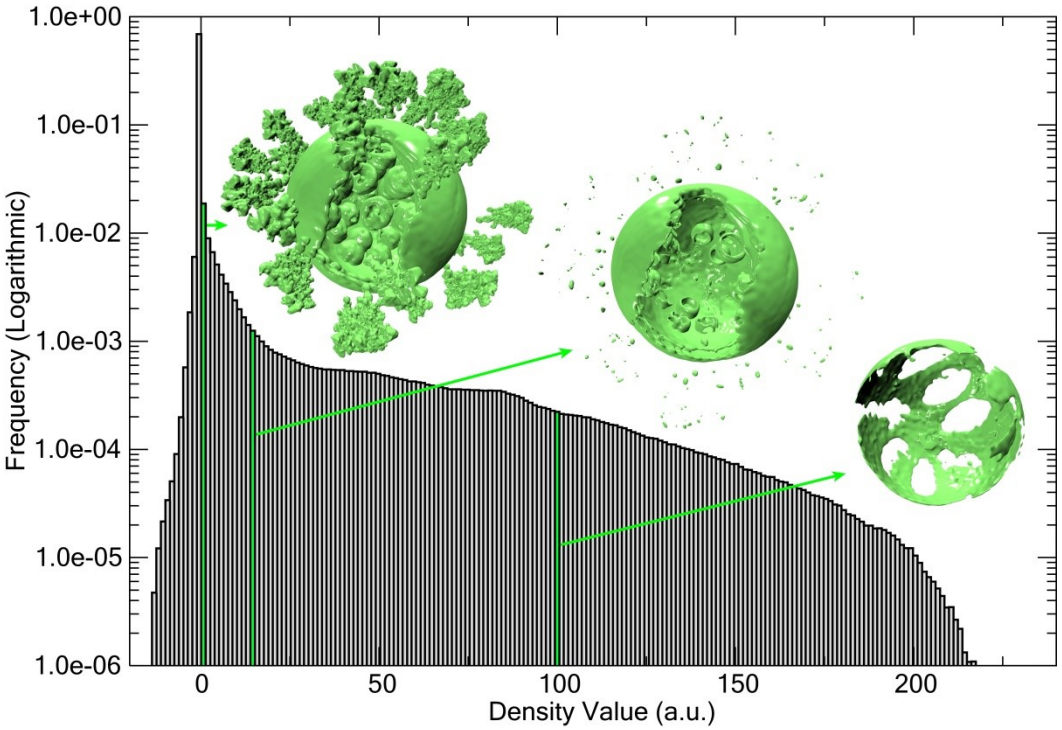

Fig. S5 Comparison of the two distinct UCG models of the SARS-CoV-2 virion. The left UCG model is developed in this work from the cryo-ET data. The right one, which was visualized from the PDB file deposited online (https://github.com/alvinyu33/sars-cov-2-public), was constructed by Yu et al. in the Voth group (*Biophys. J.,* 2021, **120,** 1097–1104).

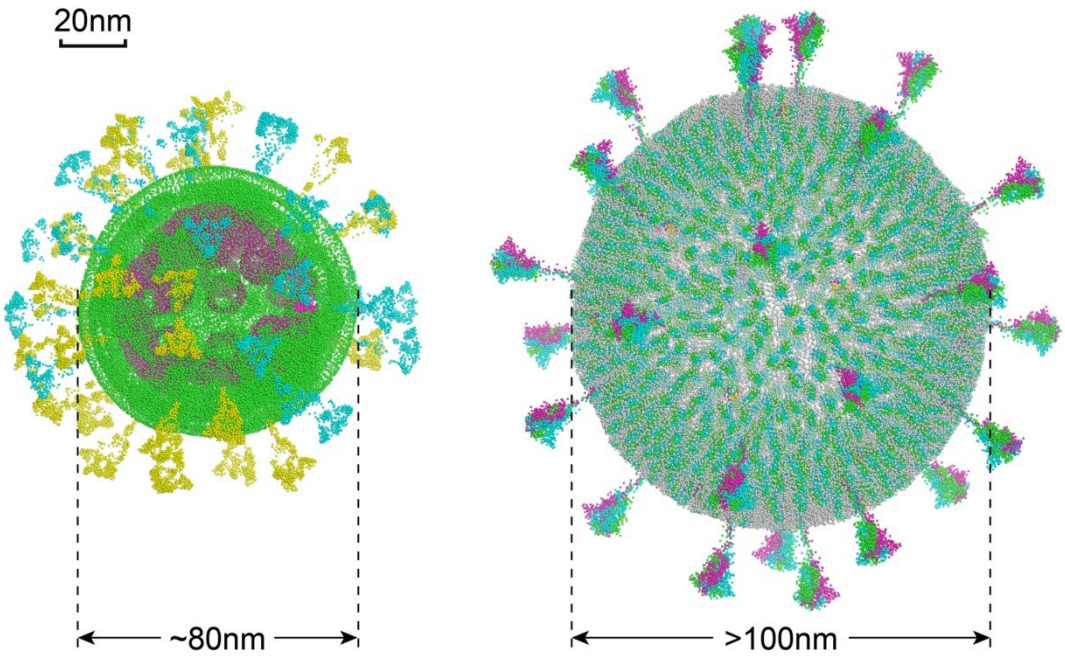

Fig. S6 The crystal structures of the closed and open conformations of the head parts of the spike proteins. The closed conformation (PDB ID: 6VXX) has three down RBDs, while the right conformation (PDB ID: 6VSB) has two down RBDs, and one up RBD shown in the red circle.

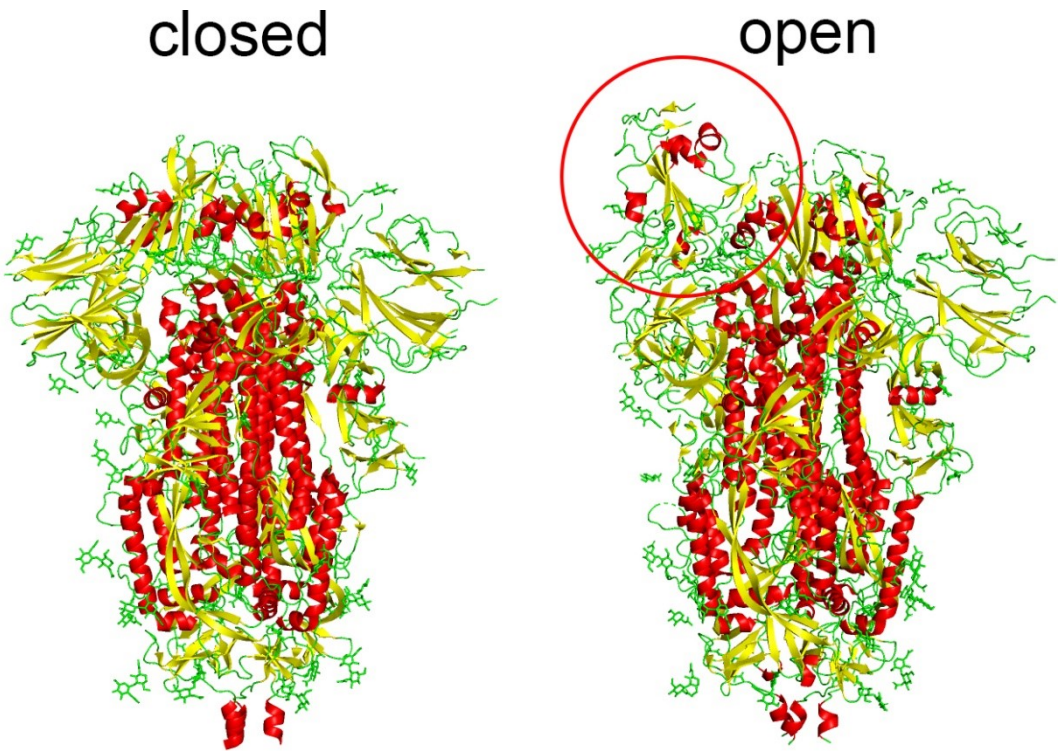

Fig. S7 Projections of the MD trajectories of 50 spike proteins from the MD trajectories of the two UCG models. The origin and blue dots belong to the UCG models with or without the RNPs, respectively.

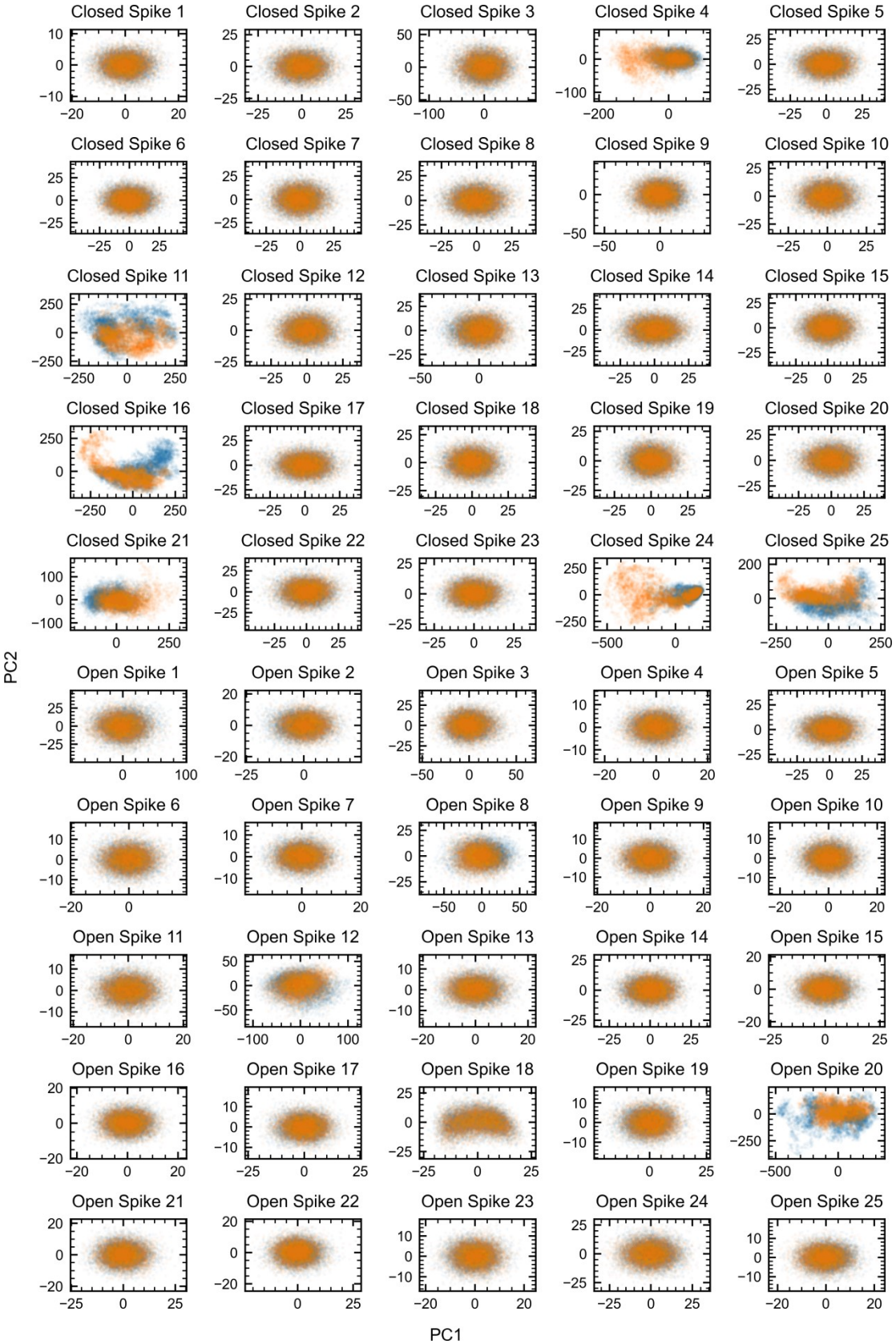

Fig. S8 The calculated RMSF curves for (a) chain B and (b) chain C of the free and the embedded spike proteins. The Cα atoms in the stalk and the CT regions are highlighted in the pink background. The MD trajectories of the RBDs in (c) chain B and (d) chain C are projected with respect to PC1 and PC2.