The Landscape-Based Protein Stability Analysis and Network Modeling of Multiple Conformational States of the SARS-CoV-2 Spike D614 Mutant: Conformational Plasticity and Frustration-Driven Allostery as Energetic Drivers of Highly Transmissible Spike Variant

Gennady Verkhivker,^{1,2*} Steve Agajanian¹, Ryan Kassab¹, Keerthi Krishnan¹

¹ Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA 92866, United States of America

² Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA 92618, United States of America

* Correspondence: verkhivk@chapman.edu; Tel.: +1-714-516-4586 (G.V)

Abstract

The structural and functional studies of the SARS-CoV-2 spike protein variants revealed an important role of the D614G mutation that is shared across many variants of concern (VOCs), suggesting the effect of this mutation on the enhanced virus infectivity and transmissibility. The recent structural and biophysical studies provided important evidence about multiple conformational substates of the D614G spike protein. The development of a plausible mechanistic model which can explain the experimental observations from a more unified thermodynamic perspective is an important objective of the current work. In this study, we employed efficient and accurate coarse-grained simulations of multiple structural substates of the D614G spike trimers together with the ensemble-based mutational frustration analysis to characterize the dynamics signatures of the conformational landscapes. By combining the local frustration profiling of the conformational states with residue-based mutational scanning of protein stability and network analysis of allosteric interactions and communications, we determine the patterns of mutational sensitivity in the functional regions and sites of variants. We found that the D614G mutation may induce a considerable conformational adaptability of the open states in the SARS-CoV-2 spike protein without compromising folding stability and integrity of the spike protein. The results suggest that the D614G mutant may employ a hinge-shift mechanism in which the dynamic couplings between the site of mutation and the inter-protomer hinge modulate the inter-domain interactions, global mobility change and the increased stability of the open form. This study proposes that mutation-induced modulation of the conformational flexibility and energetic frustration at the inter-protomer interfaces may serve as an efficient mechanism for allosteric regulation of the SARS-CoV-2 spike proteins.

Introduction

SARS-CoV-2 infection is transmitted when the viral spike (S) glycoprotein binds to the host cell receptor ACE2, leading to the entry of S protein into host cells and membrane fusion.^{1,2} The full-length SARS-CoV-2 S protein consists of amino (N)-terminal S1 subunit and carboxyl (C)-terminal S2 subunit where S1 is involved in the interactions with the host receptor and includes an N-terminal domain (NTD), the receptor-binding domain (RBD), and two structurally conserved subdomains (SD1 and SD2). The S2 subunit is highly evolutionary conserved and contains an N-terminal hydrophobic fusion peptide (FP), fusion peptide proximal region (FPPR), heptad repeat 1 (HR1), central helix region (CH), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM) and cytoplasmic tail (CT) regions (Figure S1). Structural and biochemical studies established that the mechanism of virus infection may involve conformational transitions between distinct functional forms of the SARS-CoV-2 S protein in which the RBDs continuously switch between “down” and “up” positions.³⁻¹⁰ The cryo-EM structure of the SARS-CoV-2 S trimer revealed a spectrum of closed states that included a structurally rigid closed form and more dynamic closed states preceding a transition to the fully open S conformation.⁵ Protein engineering and structural studies also showed that specific disulfide bonds and proline mutations can modulate stability of the SARS-CoV-2 S trimer⁶ and lead to the thermodynamic shifts between the closed and open forms.⁷⁻⁹ Dynamic structural changes that accompany SARS-CoV-2 S binding with the ACE2 host receptor were described in cryo-EM experiments showing a cascade of conformational transitions from a compact closed form weakened after furin cleavage to the partially open states and subsequently the ACE2-bound open form thus priming the S protein for fusion.¹⁰ The cryo-EM and tomography examined conformational flexibility and distribution of S trimers in situ on the

virion surface¹¹ showing that spontaneous conformational changes and population shifts between different functional states can be maintained in different biological environments, reflecting the intrinsic properties of conformational landscapes for SARS-CoV-2 S trimers. Single-molecule Fluorescence (Förster) Resonance Energy Transfer (smFRET) studies of SARS-CoV-2 S trimer on virus particles revealed a sequence of conformational transitions from the closed state to the receptor-bound open state suggesting that mechanisms of conformational selection and receptor-induced structural adaptation can work synchronously and showing that SARS-CoV-2 neutralization can be achieved by antibodies that either directly compete with the ACE2 receptor binding or exert their effect by allosterically stabilizing the S protein in its RBD-down conformation.¹²

The vast and continuously expanding body of structural and biochemical studies of the SARS-CoV-2 S complexes with different classes of potent antibodies targeting distinct binding epitopes of the S-RBD as well as various antibody cocktails and combinations have revealed multiple conformation-dependent epitopes, highlighting the link between conformational plasticity and adaptability of S proteins and capacity for eliciting specific binding and broad neutralization responses.¹³⁻³⁶ Deep mutagenesis scanning of antibody-escaping mutations showed that the escape mutations cluster in several RBD regions and can be constrained with respect to their effects on expression of properly folded RBD and ACE2 binding.³⁷⁻⁴² Functional and structural studies explored how B.1.1.7 (alpha), B.1.351 (beta), P.1 (gamma) and B.1.1.427/B.1.429 (epsilon) variants in the SARS-CoV-2 S protein affect conformational landscapes of the S protein and the ability to evade host immunity and incur resistance to antibodies.⁴³⁻⁵³ These studies revealed subtle structural and functional impact of mutations that

can modulate dynamics and stability of the closed and open forms, increase binding to the human receptor ACE2, and confer resistance to neutralizing antibodies.

Strikingly, all these emerging variants of SARS-CoV-2 designated into variants of concern (VOCs) share D614G mutation, that individually was linked with the enhanced infectivity profile and attracted significant early attention following the evidence of the mutation enrichment via epidemiological surveillance.^{54,55} The initial structural studies showed that the D614G mutation can act by shifting the population of the SARS-CoV-2 S trimer from the closed form (53% of the equilibrium) in the native spike protein to a widely-open topology of the “up” protomers in the D614G mutant with 36% of the population adopting a single open protomer, 39% with two open protomers and 20% with all three protomers in the open conformation.⁵⁶ The cryo-EM structures of the S-D614 and S-G614 ectodomains showed the increased population of the 1-RBD-up open form as compared to the closed state in the S-GSAS/D614G structure.⁵⁷ The electron microscopy analysis also revealed the higher 84% percentage of the 1-up RBD conformation in the S-G614 protein.⁵⁸ Functional studies showed that the S-G614 mutant exhibited the greater infectivity than the S-D614 protein which was attributed to the greater stability of the S-G614 mutant and leading to the reduced S1 subdomain shedding.⁵⁹ An alternative mechanistic scenario proposed that D614G mutation would promote rather than limit shedding of the S1 domain by altering the transitions of “up” and “down” forms.⁶⁰ The increased stability of the D614G mutant was inferred from the recent cryo-EM structures of a full-length unmodified S-G614 trimer that can reversibly adopt an all-down closed state, an intermediate (1 RBD in an intermediate state), and 1 RBD-up open conformation.⁶¹ These studies showed that the S-D614 protein was less stable, while the S-G614 mutant was stabilized in the 1 RBD-up conformations that featured ordering of the so-

called 630-loop (residues 620-640) near the interface of S1 and S2 thus helping to strengthen the inter-domain interactions and enhance the stability of the mutated S protein. This study provided support to the reduced shedding mechanism induced by the D614G mutation that inhibits a premature dissociation of the S1 subunit which eventually leads to the increased number of functional spikes and stronger infectivity.⁶¹ Structure-based protein design and cryo-EM structure determination provided an additional support to the stability hypothesis showing that D614G and D614N mutations can result in the increased stability of the S protein mutants and decrease the premature shedding of the S1 domain.⁶² The recent series of cryo-EM studies have further expanded structural characterization of the S-G614 trimer conformations, revealing a range of more flexible closed and open conformations, showing that the S-G614 occupies predominantly the open conformation with 87 % of the population in the 1 RBD-up or 2-RBD-up forms compared with 17% in the D614 structure for which 2 RBD-up form was never detected.⁶³ Cryo-EM structures of trimeric S protein in multiple states including the all RBD-locked, activated, S1/S2 trypsin-cleaved, and the full-length ACE2-bound states showed that S-G614 and the trypsin-cleaved S-G614 proteins adopt 1 RBD-up conformation confirming that D614G mutation can increase the flexibility and population of the open spike states which might contribute to the increased infectivity.⁶⁴ By combining cryo-EM analysis with differential scanning calorimetry and differential scanning fluorimetry, the recent illuminating investigation characterized multiple substates of the of the S-G614 protein in open and closed forms, showing a significant conformational heterogeneity and thermodynamic preferences of the S-G614 for 1 RBD-up and 2 RBD-up open states.⁶⁵ Using an array of biophysical tools, this study convincingly demonstrated the increased stability of the S-G614 trimer manifested in the increased resistance to cold- and heat-induced unfolding, and

suggested that the increased conformational flexibility and emergence of multiple substates of the open form may be the main thermodynamic driver promoting virus infectivity.⁶⁵ These findings were echoed in another biophysical study showing that the S-G614 variant was more stable than the S-D614 variant and had better binding ability with the ACE2 receptor after storage at -20°C for up to 30 days.⁶⁶ This study supported the notion that the increased stability and infectivity of the S-G614 variant are intimately connected and the higher stability of S-G614 compared to that of S-D614 may contribute to rapid viral spread.⁶⁶ The biophysical experiments using bilayer interferometry technology revealed that the S-G614 variant displays highest entry efficiency among natural S variants and the increased binding affinity for the ACE2 protein dimer as compared to S-D614 protein in a temperature-dependent manner, suggesting that mutation-induced conformational flexibility of the S protein may be responsible for the observed ~2.0 -fold increase in the binding affinity.⁶⁷

The multiscale structural modeling and atomistic molecular dynamics (MD) simulations of the full-length SARS-CoV-2 S glycoprotein embedded in the viral membrane, with a complete glycosylation profile provided the blueprint for biophysical modeling of S proteins.⁶⁸⁻⁷⁴ Using deep data analysis and protein structure network modeling of MD simulations, another computational study identified residues that exhibit long-distance coupling with the RBD opening, including sites harboring functional mutations D614G and A570D which points to the important role of D614G variant in modulating allosteric communications in the S protein.⁷⁵ The free energy landscapes of the S protein derived from MD simulations together with nudged elastic pathway optimization mapping of the RBD opening revealed a specific transient allosteric pocket at the hinge region that is located near D614 position influences RBD dynamics.⁷⁶ Using computational modeling it was suggested that the D614G mutation may

affect the inter-protomer energetics and induce specific interaction changes in the C-terminal domain 1(CTD1) (residues 528-591) and fusion peptide (FP) regions that mediate couplings between S1 and S2 subunits.⁷⁷ MD simulations probed the effects of the D614G mutation, suggesting that the variant favors an open conformation in which S-G614 protein maintains the symmetry in the number of persistent contacts between the three protomers.⁷⁸ Markov modeling characterized the dynamics of the S protein and mutational variants, predicting the increase in the open state occupancy for the D614G mutation due to the increased flexibility of the closed state and the enhanced stabilization of the open form.⁷⁹ The energy analysis of the S-D614 and S-G614 proteins in the closed and partially open conformations showed that local interactions near D614 position may be energetically frustrated and become stabilized in the S-G614 mutant through strengthening of the inter-protomer association between S1 and S2 regions.⁸⁰ Structure-based model showed that the D614G mutation may promote closer association and stronger interactions with S2 subunit, supporting the reduced shedding mechanism of the increased S-614 stability.⁸¹ Molecular simulations and network modeling approaches were used on our most recent investigation to present evidence that the SARS-CoV-2 spike protein can function as an allosteric regulatory engine that fluctuates between dynamically distinct functional states.^{82,83} Allosteric models of the SARS-CoV-2 S protein dynamics proposed in our studies have allowed for better understanding distant couplings between S protein regions and detection of allosteric hotspot switches regulating binding with antibodies and nanobodies.⁸⁴⁻⁸⁹

The existing mechanisms suggest that mutation-induced allosteric transitions to the open states and the reduced shedding mechanism of the thermodynamic stability of the open S-G614 trimers are important driving forces for greater infectivity of the D614G mutant. However, a

consensus view on the mechanism underlying the functional effects and increased infectivity of S-D614G spike mutant is yet to be established. The development of a plausible mechanistic model which can explain the experimental observations from a unified dynamic perspective that considers local effects and long-range allosteric interactions is an important objective of current investigations. In this study, we employed the efficient and accurate coarse-grained simulations of multiple structural substates of the S-G614 trimers to characterize the dynamics signatures of the conformational landscapes of spike proteins. By combining atomistic simulations with the ensemble-based local frustration analysis of the conformational states, we determine sensitivity of functional regions and sites of mutational variants to energetic variations in different functional states. Using the ensemble-based local frustration profiling and mutational scanning of protein stability changes in the S-G614 conformational states, we quantify conformational plasticity and adaptability of the S-G614 states, showing that the inter-protomer interactions can be generally weakened in significantly heterogeneous and expanded S-G614 open states. Combined, the ensemble-based dynamics and energetic analysis of multiple S-G614 substates suggests a significant conformational and mutational plasticity of the open S-G614 trimer. We show that disorder-order changes in the flexible loop near G614 may be coupled with the energetic frustration in the mutational site and allow for modulation of the inter-protomer interactions. This study suggests that modulation of the energetic frustration at the inter-protomer interfaces by D614G mutation can serve as a mechanism for allosteric regulation in which the dynamic couplings between the site of mutation and the inter-protomer hinge of functional motions would modulate the inter-domain interactions, global changes in mobility and the increased stability of the open form.

Materials and Methods

Structure Preparation and Analysis

All structures were obtained from the Protein Data Bank.^{90,91} During structure preparation stage, protein residues in the crystal structures were inspected for missing residues and protons. Hydrogen atoms and missing residues were initially added and assigned according to the WHATIF program web interface.^{92,93} The structures were further pre-processed through the Protein Preparation Wizard (Schrödinger, LLC, New York, NY) and included the check of bond order, assignment and adjustment of ionization states, formation of disulphide bonds, removal of crystallographic water molecules and co-factors, capping of the termini, assignment of partial charges, and addition of possible missing atoms and side chains that were not assigned in the initial processing with the WHATIF program. The missing loops in the studied cryo-EM structures of the SARS-CoV-2 S protein were reconstructed and optimized using template-based loop prediction approaches ModLoop⁹⁴, ArchPRED Server⁹⁵ and further confirmed by FALC (Fragment Assembly and Loop Closure) program.⁹⁶ The side chain rotamers were refined and optimized by SCWRL4 tool.⁹⁷ The conformational ensembles were also subjected to all-atom reconstruction using PULCHRA method⁹⁸ and CG2AA tool⁹⁹ to produce atomistic models of simulation trajectories. The protein structures were then optimized using atomic-level energy minimization with a composite physics and knowledge-based force fields as implemented in the 3Drefine method.¹⁰⁰ The atomistic structures from simulation trajectories were further elaborated by adding N-acetyl glycosamine (NAG) glycan residues and optimized. The glycosylated microenvironment for atomistic models of the simulation trajectories was mimicked by using the structurally resolved glycan conformations for 22 most occupied N-glycans^{101,102} in each as determined in the cryo-EM structures of the SARS-CoV-2 spike S

trimer in the closed state (K986P/V987P,) (pdb id 6VXX) and open state (pdb id 6VYB), and the cryo-EM structure SARS-CoV-2 spike trimer (K986P/V987P) in the open state (pdb id 6VSB).

Coarse-Grained Simulations and Elastic Models

Coarse-grained (CG) models are computationally effective approaches for simulations of large systems over long timescales. We employed CABS-flex approach that efficiently combines a high-resolution coarse-grained model and efficient search protocol capable of accurately reproducing all-atom MD simulation trajectories and dynamic profiles of large biomolecules on a long time scale.¹⁰³⁻¹⁰⁹ In the CG-CABS model, the amino acid residues are represented by $C\alpha$, $C\beta$, the center of mass of side chains and another pseudoatom placed in the center of the $C\alpha$ - $C\alpha$ pseudo-bond. The sampling scheme of the CABS model used in our study is based on Monte Carlo replica-exchange dynamics and involves a sequence of local moves of individual amino acids in the protein structure as well as moves of small fragments.¹⁰³⁻¹⁰⁵ CABS-flex standalone package dynamics implemented as a Python 2.7 object-oriented package was used for fast simulations of protein structures.¹⁰⁷⁻¹⁰⁹ In the CABS-flex package we also used MODELLER-based reconstruction of generated models and simulation trajectories to all-atom representation. The default settings were applied in which soft native-like restraints are imposed only on pairs of residues fulfilling the following conditions : the distance between their C^α atoms was smaller than 8 Å, and both residues belong to the same secondary structure elements. A total of 1,000 independent CG-CABS simulations were performed for each of the studied systems. In each simulation, the total number of cycles was set to 10,000 and the number of cycles between trajectory frames was 100.

We performed principal component analysis (PCA) of reconstructed trajectories derived from CABS-CG simulations using the CARMA package¹¹⁰ and also determined the essential dynamics profiles in slow modes using elastic network models (ENM) analysis.¹¹¹ Two elastic network models: Gaussian network model (GNM)^{112,113} and Anisotropic network model (ANM) approaches¹¹⁴ were adopted to compute the amplitudes of isotropic thermal motions and directionality of anisotropic motions. The functional dynamics analysis was conducted using the GNM in which protein structure is reduced to a network of N residue nodes identified by C_α atoms and the fluctuations of each node are assumed to be isotropic and Gaussian. The topology of the protein structure is described by $N \times N$ Kirchhoff matrix of inter-residue contacts Γ , where the off-diagonal elements are -1 , if the nodes are within a cutoff distance r_c , and zero otherwise. In GNM approach the interaction potential for a protein of N residues was:

$$V_{\text{GNM}} = -\frac{\gamma}{2} \left[\sum_{i=1}^{N-1} \sum_{j=i+1}^N (R_{ij} - R_{ij}^0) \cdot (R_{ij} - R_{ij}^0) \Gamma_{ij} \right] \quad (1)$$

where R_{ij}^0 and R_{ij} were the equilibrium and instantaneous distance between residues i and j , and Γ was the $N \times N$ Kirchhoff matrix:

$$\Gamma_{ij} = \begin{cases} -1 & i \neq j, R_{ij} \leq r_c \\ 0 & i \neq j, R_{ij} > r_c \\ -\sum_{i, i \neq j} \Gamma_{ij} & i = j \end{cases} \quad (2)$$

Thus, the square fluctuations were:

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \left(3kT/\gamma \right) \cdot [\Gamma^{-1}]_{ij} \quad (3)$$

The normal modes were extracted by eigenvalue decomposition of the matrix $\Gamma = U\Lambda U^T$, U being the orthogonal matrix whose k^{th} column u_k was the k^{th} mode eigenvector. Λ was the diagonal matrix of eigenvalues λ_k .

In ANM approach⁹² the interaction potential for a protein of N residues was:

$$V_{\text{ANM}} = \frac{\gamma}{2} \sum_{i,j}^N (|R_{ij}| - |R_{ij}^0|)^2 \quad (4)$$

The $3N \times 3N$ Hessian matrix, H , determines the systems dynamics.

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \quad (5)$$

where X_i , Y_i and Z_i represented the Cartesian components of the i -th residue, V was the potential energy of the system. We selected $r_c = 13 \text{ \AA}$. ANMs provide information about the amplitudes, as well as the direction of residue fluctuations. Conformational mobility profiles in the essential space of low frequency modes were obtained using DynOmics server¹¹³ and ANM server.¹¹⁴

Mutational Sensitivity Analysis and Alanine Scanning

To compute protein stability changes in the SARS-CoV-2 S structures, we conducted a systematic alanine scanning of protein residues in the SARS-CoV-2 trimer mutants as well as mutational sensitivity analysis at the mutational site for both SARS-CoV-2 S-D614 and SARS-CoV-2 S-G614 structures. Two different approaches were used. Alanine scanning of protein residues was performed using FoldX approach.¹¹⁵⁻¹²⁰ and BeAtMuSiC approach.¹²¹⁻¹²³ If a free energy change between a mutant and the wild type (WT) proteins $\Delta\Delta G = \Delta G(\text{MT}) - \Delta G(\text{WT}) > 0$, the mutation is destabilizing, while when $\Delta\Delta G < 0$ the respective mutation is stabilizing.

BeAtMuSiC approach is based on statistical potentials describing the pairwise inter-residue distances, backbone torsion angles and solvent accessibilities, and considers the effect of the mutation on the strength of the interactions at the interface and on the overall stability of the complex.¹²¹⁻¹²³ We leveraged rapid calculations based on statistical potentials to compute the ensemble-averaged alanine scanning computations and mutational sensitivity analysis at D614 and G614 positions using equilibrium samples from reconstructed simulation trajectories.

Dynamic-Based Modeling of Residue Interaction Network and Community Analysis

A graph-based representation of protein structures^{124,125} is used to represent residues as network nodes and the inter-residue edges to describe non-covalent residue interactions. The network edges that define residue connectivity are based on non-covalent interactions between residue side-chains that define the interaction strength I_{ij} according to the following expression used in the original studies.^{124,125}

$$I_{ij} = \frac{n_{ij}}{\sqrt{(N_i \times N_j)}} \times 100 \quad (6)$$

where n_{ij} is number of distinct atom pairs between the side chains of amino acid residues i and j that lie within a distance of 4.5 Å. N_i and N_j are the normalization factors for residues i and j . We constructed the residue interaction networks using both dynamic correlations¹²⁶ and coevolutionary residue couplings¹²⁷ that yield robust network signatures of long-range couplings and communications. The details of this model were described in our previous studies.^{128,129} In brief, the edges in the residue interaction network are then weighted based on dynamic residue correlations and coevolutionary couplings measured by the mutual information scores.

The edge lengths in the network are obtained using the generalized correlation coefficients $R_{MI}(X_i, X_j)$ associated with the dynamic correlation and mutual information shared by each pair of residues. The length (i.e. weight) $w_{ij} = -\log[R_{MI}(X_i, X_j)]$ of the edge that connects nodes i and j is defined as the element of a matrix measuring the generalized correlation coefficient $R_{MI}(X_i, X_j)$ as between residue fluctuations in structural and coevolutionary dimensions. Network edges were weighted for residue pairs with $R_{MI}(X_i, X_j) > 0.5$ in at least one independent simulation as was described in our initial study. The matrix of communication distances is obtained using generalized correlation between composite variables describing both dynamic positions of residues and coevolutionary mutual information between residues. As a result, the weighted graph model defines a residue interaction network that favors a global flow of information through edges between residues associated with dynamics correlations and coevolutionary dependencies.

To characterize allosteric couplings of the protein residues and account for cumulative effect of dynamic and coevolutionary correlations, we employed the generalized correlation coefficient first proposed by Lange and Grubmüller.¹³⁰ The `g_correlation` tool in the Gromacs 3.3 package was used that allows computation of both linear or non-linear generalized correlation coefficients.¹³¹ This approach has also been utilized in a similar context of allosteric modeling by other groups¹³²⁻¹³⁵ where the generalized correlation GC matrix measured dynamic interdependencies between on fluctuations of spatially separated residues.

The RING program^{136,137} was also employed for the initial generation of residue interaction networks. The ensemble of shortest paths is determined from matrix of communication distances by the Floyd-Warshall algorithm.¹³⁸ Network graph calculations were performed using the python package NetworkX.¹³⁹

The Girvan-Newman algorithm¹⁴⁰⁻¹⁴² is used to identify local communities. In this approach, edge centrality (also termed as edge betweenness) is defined as the ratio of all the shortest paths passing through a particular edge to the total number of shortest paths in the network. The method employs an iterative elimination of edges with the highest number of the shortest paths that go through them. By eliminating edges, the network breaks down into smaller communities. The algorithm starts with one vertex, calculates edge weights for paths going through that vertex, and then repeats it for every vertex in the graph and sums the weights for every edge. However, in complex and dynamic protein structure networks it is often that number of edges could have the same highest edge betweenness.

An improvement of Girvan-Newman method was implemented, and the algorithmic details of this modified scheme were given in our recent studies.¹⁴³⁻¹⁴⁵ Briefly, in this modification of Girvan-Newman method, instead of a single highest edge betweenness removal, all highest betweenness edges are removed at each step of the protocol. This modification makes community structure determination invariant to the labeling of the nodes in the graph and leads to a more stable solution. The modified algorithm proceeds through the following steps : a) Calculate edge betweenness for every edge in the graph; b) Remove all edges with highest edge betweenness within a given threshold; c) Recalculate edge betweenness for remaining edges; d) Repeat steps b-d until graph is empty.

The betweenness of residue i is defined as the sum of the fraction of shortest paths between all pairs of residues that pass through residue i :

$$B_k = \sum_{s \neq k \neq t}^N \frac{S_{st}(k)}{S_{st}} \quad (7)$$

where S_{st} denotes the number of shortest geodesics paths connecting s and t , and $S_{st}(k)$ is the number of shortest paths between residues s and t passing through the node k .

The following Z-score is then calculated

$$Z_i = \frac{B_k - \langle B \rangle}{\sigma} \quad (8)$$

Through mutation-based perturbations of protein residues we compute changes in the average short path length (ASPL) averaged over all possible modifications in a given position.

The change of ASPL upon mutational changes of each node is given as

$$\Delta L_i = \langle ||\Delta L_i^{node}(j)||^2 \rangle \quad (9)$$

where i is a given site, j is a mutation and $\langle \dots \rangle$ denotes averaging over mutations. $\Delta L_i^{node}(j)$

describes the change of ASPL upon mutation j in a residue node i . ΔL_i is the average change of ASPL triggered by mutational profiling of this position.

Z-score is then calculated for each node as follows:

$$Z_i = \frac{\Delta L_i - \langle \Delta L \rangle}{\sigma} \quad (10)$$

The ensemble-averaged Z-scores ASPL changes are computed from network analysis of the conformational ensembles using 1,000 snapshots of the simulation trajectory for the native protein system.

Results and Discussion

Conformational Mobility Profiles of the SARS-CoV-2 S-G614 Proteins: The D614G Mutation Induces Differential Conformational Plasticity in Multiple Conformational Substates

We employed multiple CG simulations followed by atomistic reconstruction and refinement to provide a comparative analysis of the dynamic landscapes characteristic of the major functional states of the SARS-CoV-2 S trimer. While all-atom MD simulations with the explicit inclusion of the glycosylation shield could provide a rigorous assessment of conformational landscape of the SARS-CoV-2 S proteins, such direct simulations remain to be technically challenging due to the size of a complete SARS-CoV-2 S system embedded onto the membrane. To partially address these computational challenges, and enable analysis of multiple conformational states, we combined CG simulations with atomistic reconstruction and additional optimization by adding the glycosylated microenvironment as described in Materials and Methods. The presented analysis characterized the intrinsic dynamics of the SARS-CoV-2 S proteins in the D614G mutant states, particularly probing the diversity of the open S-D614G conformations (Figures 1,2).

We first performed simulations of the S-G614 mutant structures in the locked closed, intermediate and open forms (pdb id 7KRQ, 7KRS, and 7KRR respectively) solved by Chen and colleagues.⁶¹ These cryo-EM structures of the full-length S protein in the prefusion conformation discovered an ordered conformation of ~25-residue FPPR segment that is typically disordered in other structures, demonstrating that RBDs can be locked in the tightly packed and rigid closed form (Figure 1A, D). The reported S-G614 mutant structures revealed structurally ordered 630 loop (residues 617 to 644) and FPPR region (residues 823 to 862). In the closed

form of the S-G614 mutant, the FPPR segments (residues 823-862) and 630 loops are fully ordered (Figure 1A,D). In the intermediate conformation with a partly open single RBD, the 630 loops are disordered in the RBD-shifted module and FPPR remain fully ordered (Figure 1B, E).

Figure 1. Cryo-EM structures of the SARS-CoV-2 S-G614 trimer structures and composition of the used in this study. The cryo-EM structures of SARS-CoV-2 S-G614 in the closed state, pdb id 7KRQ (A), in the intermediate state, pdb id 7KRS (B) and 1 RBD-up open form, pdb id 7KRR (C). The structures are shown in full spheres and colored with protomers A,B,C are colored in green, red and blue. The position of G614 is shown in spheres. (D-F) Structures of the 630 loop (residues 617 to 644) and FPPR (residues 823 to 862) are shown for protomer A in green ribbons, protomer B in red ribbons and protomer C in blue ribbons. The rendering of SARS-CoV-2 S structures was done using the interactive visualization program UCSF ChimeraX package.¹⁴⁶

Conformational dynamics analysis showed a considerable difference in the mobility profiles for these three conformational states of the S-G614 mutant (Figure 2A). Indeed, we found that the closed S-G614 trimer was generally stable, showing moderate but widely-spread fluctuations in the NTD and RBD regions (Figure 2B). Intriguingly, for the intermediate S-G614 mutant form, there is only marginal increase in the mobility of the S1 subunit, particularly RBD regions, while the S2 regions become more stable (Figure 2A,C). We observed that the intermediate form may be characterized by tight packing in the S2 regions and softening in the S1 subunit, which may be relevant for consolidation of the inter-protomer hinges typically formed at the borders of highly stable and more flexible regions (Figure 2C). Strikingly, a comparative analysis of structural maps of conformational mobility highlighted these changes, showing the increased stabilization of the partially open S-G614 trimer (Figure 2B-D). For the 1 RBD-up form of the S-G614 trimer (Figure 1C, F), we detected that only RBD and NTD regions remain fairly mobile (Figure 2A, D). These results are consistent with the notion that D614G mutation may induce the greater stability of the open conformations in which the remaining two closed RBDs are locked in their respective positions by the ordered 630 loops. Hence, the conformational dynamics analysis showed that the global mobility pattern is greatly influenced by structural constraints provided by the ordered 630 loop and FPPR segment. By highlighting the mutational D614G site, we noticed that this position may be located near the inter-domain hinges that form conformationally stable anchor regions that control global movements of the highly mobile RBD and NTD domains.

We also investigated conformational dynamics of another group of the closed (pdb id 7BNM), 1 RBD-up open (pdb id 7BNN) and 2 RBD-up (pdb id 7BNO) conformations of the S-G614 protein. The construct for the D614G mutant was based on the furin-uncleavable version of the

SARS-CoV-2 spike protein ectodomain with a set of stabilizing mutations (R682S, R685S, K986P, and K987P).⁶³ Notably, in these S-G614 trimer conformations the 630 loop, the FPPR region (residues 828-853) and the furin cleavage site were disordered. It can be seen that all functional forms of the S-G614 mutant were fairly mobile, including the closed conformation (Figure 2E). In the closed state, the NTD regions, RBD (residues 331-528) and CTD1 (residues 528-591) corresponded to dynamic regions in the S1 subunit. While we observed the progressively increased fluctuations of the RBD residues in the 1 RBD-up and 2 RBD-up forms, the conformational mobility in the closed state was also evident (Figure 2E). The inter-domain regions and interfacial CTD1 residues linking S1 and S2 subunits showed an appreciable mobility level. Structural maps of conformational mobility displayed significant softening near the inter-protomer regions and at the S1/S2 interfaces (Figure 2F-H). We observed that G614 and the surrounding region displayed the increased conformational variability which may reflect structural weakening of the inter-protomer contacts and a wider topology of the S-G614 structures. These dynamic features may facilitate conformational transitions from the closed to a partially and fully open forms of S-G614 protein. The increased mobility of the S-G614 trimer in the fully open form could be seen from a comparison of structural mobility maps of the S-G614 structures (Figure 2F-H). It appeared that in the fully open form, S-614 trimer can experience larger fluctuations and partial expansions not only in the S1 subunit but also in the S2 regions that become more plastic and amenable for large functional changes. Hence, our results are consistent with structural studies⁶³ showing that S-G614 mutant can assume a wide range of conformations, particularly increasing plasticity of the closed form, which may facilitate subsequent structural rearrangements to the open form.

Figure 2. CG-CABS conformational dynamics profiles of the SARS-CoV-2 S-G614 structures. The S1 regions include NTD (residues 14-306), RBD (residues 331-528), CTD1 (residues 528-591), CTD2 (residues 592-686), UH (residues 736-781), HR1 (residues 910-985), CH (residues 986-1035), CD (residues 1035-1068), HR2 (residues 1069-1163). (A) The root mean square fluctuations (RMSF) profiles from simulations of the cryo-EM structures of SARS-CoV-2 S-G614 in the closed state, pdb id 7KRQ (green), in the intermediate state, pdb id 7KRS (red) and 1 RBD-up open form, pdb id 7KRR (blue). (B-D) Structural maps of the conformational profiles for the S-G614 closed state, the intermediate state and 1 RBD-up open form respectively. (E) The RMSF profiles of the cryo-EM structures of SARS-CoV-2 S-G614 in the closed state, pdb id 7BNM (green), 1 RBD-up state, pdb id 7BNN (red) and 2 RBD-up open form, pdb id 7BNO (blue). (F-H) Structural maps of the conformational mobility profiles for these S-G614 structures. The position of mutational site D614G is shown in spheres. The structures are in sphere-based representation rendered using UCSF ChimeraX package¹⁴⁶ with the rigidity-to-flexibility sliding scale colored from blue to red.

We also simulated a group of the S-G614 open structures that represent three distinct 1 RBD-up substrates and two different 2 RBD-up states characterized by different angle of the upward rotation with respect to the hinge region.⁶⁵ These structures did not show significant conformational rearrangement around the interprotomer interaction site between G614 and T859 but revealed partially disordered K854 loop⁶⁵ (Figure 3).

Structural maps of conformational mobility profiles obtained from CG-CABS simulations of these states revealed a fairly rigid S2 subdomain, mobile NTD and RBD regions for the 1-up protomer and the appreciable flexibility for one of the closed-down protomers (Figure 3A-C). This may reflect the increased propensity of 1 RBD-up conformations priming one of the closed protomers for the transition to the open state. The increased population of the RBD-up conformations and plasticity of the multiple open states may allow for specific modulation of the RBD motions and greater accessibility to the ACE2 host receptor.

The obtained conformational mobility profile for multiple 1 RBD-up substates is relatively similar, showing the rigidity of the S2 regions and mobility of the RBD-up regions (Figure 3A-C). It was also observed that the region harboring G614 site can bridge rigid and more flexible regions at the S1/S2 interface, with the site of mutation residing in stable local structural environment. These results suggest that modulation of local flexibility near the G614 position may affect the interfacial hinge regions and global movements of the RBD regions. Structural maps of the mobility profiles for 2 RBD-up states highlighted the increased plasticity of the S-G614 conformations in which both S1 and S2 subunits become “softer” and may experience significant fluctuations at the inter-protomer and S1/S2 interfacial regions (Figure 3D,E).

Figure 3. Structural maps of the conformational mobility profiles for the SARS-CoV-2 S-G614 1 RBD-up states (pdb id 7EAZ, 7EB0, 7EB3) (A-C) and 2 RBD-up states (pdb id 7EB4, 7EB5) (D,E). (E) The position of mutational site D614G is highlighted in spheres and annotated. The structures are in sphere-based representation rendered using UCSF ChimeraX package¹⁴⁶ with the rigidity-to-flexibility sliding scale colored from blue to red.

As a result, the hinge regions could migrate away from their positions, making the site of mutation and its local environment more flexible and decoupled from motions in the RBD regions. Hinges are the structurally rigid regions that link higher flexibility regions and may control their relative functional motions. Previous computational studies suggested that an enzyme evolves with subtle changes in dynamics and function, through a series of hinge-shift mutations by altering distinct, flexible positions with rigid positions. Our analysis indicated that changes in the inter-protomer hinge regions in the 2 RBD-up conformations and the increased

conformational plasticity in the S1 subunit can further increase the diversity of receptor-accessible RBD conformations and promote association with the host receptor. Given that all considered structures may represent a spectrum of available conformational states for the S-G614 mutant, we infer that the S-G614 protein may sample a range of conformational states that promote frequent transitions to the open form.

Functional Slow Modes of the SARS-CoV-2S-G614 Conformations Reveal Mediating Role of G614 in Communication of Hinge Regions Controlling Transitions between Open and Closed States

To identify hinge sites and characterize collective motions in the SARS-CoV-2 S-D614 and SARS-CoV-2 S-G614 structures, we performed PCA of atomistic reconstructed trajectories derived from CG simulations. Overall, the key functional signature of collective dynamics in these states are the preferences for NTD and RBD motions, suggesting that the conformational dynamics of these states can enable functional movements of RBDs to the erected, receptor-accessible conformation (Figure 4). Although functional slow modes follow a generally similar pattern for other group of S-G614 states, there are important differences reflecting the stability of the closed and open forms. The conserved hinge regions in the closed and open forms of the S-D614 protein that can regulate the inter-domain movements between RBD and NTD as well as the relative motions of S1 and S2 regions. We first compared the slow mode profile averaged over the first three lowest mode for the constrained S-G614 conformations (Figure S2). The major hinge positions are located at F318, F592, A570, I572, Q613, G614 and Y855 residues. These sites are specified in the distributions of slow modes (Figure S2) and emerged as

immobilized islands in both the closed and open states. Notably, A570 and F592 hinge residues are situated near the inter-domain SD1-S2 interfaces and could act collectively as regulatory switch centers governing the population shifts between closed and open forms. Notably, for both 1 RBD-up and 2 RBD-up open conformations we observed significant functional displacements of at least two RBDs which suggests a significant degree of functional plasticity of the RBDs in the open conformations with the ordered 630 loop and FPPR region (Figure S2).

The observed pattern of collective motions in the S-G614 mutant would imply that the RBD regions may experience functional displacements in the intermediate form that encourage movements toward the open form. At the same time, once the S-G614 trimer attains the 1-up open form, the closed protomers become rigidified with their RBD movements being largely curtailed (Figure S2). As a result, unless open RBD becomes fully engaged by the ACE2 receptor, this may arguably prompt the open protomer to return to the down conformation as was proposed in structural studies.⁶¹

A generally similar pattern was seen in the slow mode analysis of the multiple S-G614 open conformations (Figure 4). The analysis of slow modes for the flexible open conformations⁶³ (Figure 4A,B) showed significant displacements of NTD and RBD regions. In particular, we noticed that RBD motions in the 2 RBD-up state become more pronounced. The slow mode profiles and structural maps of functional movements revealed that G614 is situated near major hinge clusters separating stable and flexible regions and can be involved in orchestrating RBD movements (Figure 4C,D).

Figure 4. Functional dynamics of the SARS-CoV-2 S S-G614 trimer structures in the open form. The essential mobility profiles are averaged over the first three major low frequency modes. The essential mobility profiles for the SARS-CoV-2 S-G614 in the 1 RBD-up form, pdb id 7BNN (A), 2 RBD-up conformation, pdb id 7BNO (B). The profiles for protomer chains A, B and C are shown in green, red and blue lines respectively. The positions of the hinge sites forming the inter-protomer cluster F318, A570, F592, G614, M740, Y855, and T859 are shown along the profiles in filled orange-colored circles. Structural maps of the essential mobility profiles for the 1 RBD-up form, pdb id 7BNN (C) and 2 RBD-up conformation, pdb id 7BNO (D). The essential mobility profiles for the SARS-CoV-2 S-G614 in the 1 RBD-up form, pdb id 7EB0 (E), 2 RBD-up conformation, pdb id 7EB5 (F). Structural maps of the essential mobility profiles for the 1 RBD-up form, pdb id 7EB0 (G) and 2 RBD-up conformation, pdb id 7EB5 (H). The structures are shown in ribbons with the rigidity-to-flexibility scale colored from blue to red.

Indeed, the mutational site is located in the immobilized structural region of the SD2 domain that together with β strand formed by residues 311–319 may correspond to a hinge center governing motions of NTD and RBD, as well motions in S1 from the S2 subunit. The flexible 28-residue N2R linker (residues 306–334) connects the NTD (residues 27–305) and RBD (residues 335–521) and stacks against the SD1 (residues 529–591) and SD2 (residues 592–697). The residues 311–319 of the linker that associate with the SD2 subdomain remained immobilized in slow modes forms a hinge region separating the mobile NTD and RBD, as well as isolating the motions in S1 from the S2 subunit. The broad hinge regions are formed by residue clusters K310-F318 that mediates NTD-RBD motions and F592-V618 that control RBD-S2 motions (Figure 4).

The slow mode profiles obtained multiple 1 RBD-up and 2 RBD-up substates of S-G614 structures⁶⁵ have a similar shape, featuring large displacements of NTD and RBD regions for the ACE2-accessible protomers (Figure 4 E,F). Of notice are appreciable displacements for the RBD-down domain in the 2 RBD-up structure (Figure 4F), particularly pointing to a peak aligned with E484 residue in this protomer. This indicated that the highly flexible 2 RBD-up states could promote the increased mobility of the closed RBD protomer and prime S-G614 for subsequent transition to 3 RBD-up state that maximizes the exposure of the S protein to binding with ACE2. Structural maps for these states highlighted the broadened area of functional motions beyond highly mobile NTD and RBD regions (Figure 4 G,H). In addition, we noticed a somewhat weakened level of immobilization for the G614 position, indicating partial reorganization of the hinge clusters in these states. These results suggested that the D614G mutation confers the increased structural flexibility in the closed state which may promote the exchange between the open and closed forms and the increased exposure of the RBDs for

interactions with ACE2. In addition to major hinge points centered on F318 and A570/F592, a more detailed analysis of the slow mode profiles also revealed presence of local minima corresponding to residues D737/M740 and K854/F855 (Figure 4). This is evident for the functional motion profiles of the constrained S-G614 states and 2 RBD-up conformations (Figure 4B,F). Hence, conformational changes in the S protein may be orchestrated through a cross-talk of several hinge centers that could couple NTD-RBD and RBD-S2 motions.

Intriguingly, structural analysis of the ensemble-averaged S-G614 states revealed an important role of G614 site in mediating communication between these hinge clusters (Figure 5). The cluster of hinge sites formed by K854, F855, F592, P589, D737 and M740 is fairly well-packed and remains intact in both closed and open S-G614 states (Figure 5). In the ensemble of structurally constrained S-G614 conformations with the ordered 630 loop, G614 is weakly linked with the hinge cluster through the key hinge position F592 (Figure 5A,B). The presence of structurally ordered loop may limit lateral movements near G614 position and provide only suboptimal interactions with F592. Nonetheless, the strategic position of G614 could allow for communication between the hinge cluster (K854, F855, F592, P589, D737 and M740) and other hinge center (F592-V618) that control S1-S2 motions. In more flexible open S-G614 conformations with disordered 630 loop, G614 can experience more significant lateral displacements by approaching and departing from F592-centered hinge cluster (Figure 5C,D). As a result, we argue that G614 site may serve as a versatile regulatory switch point that links or decouples the network of hinge clusters. In this mechanism, the disorder-order transitions of the flexible 630 loop could allosterically modulate G614 couplings with the hinge clusters and therefore contribute to the control over collective motions of the S1 subunit.

Figure 5. Structural analysis of the ensemble-averaged S-G614 states in the closed and open forms. (A) A close-up of the interaction cluster between G614 and hinge region formed by F592 and P589 of the same protomer (in green) and K835, K854, F855, M740, D737 and F855 of another protomer (in cyan) in the closed form (pdb id 7KRQ). (B) A close-up of the interaction cluster between G614 and hinge region in the 1 RBD-up open form (pdb id 7KRR). (C) A close-up of the interaction cluster between G614 and hinge region in the 1 RBD-up open form (pdb id 7BNN). (D) A close-up of the interaction cluster between G614 and hinge region in the 1 RBD-up open form (pdb id 7EB0). Note that G614 is linked to the hinge cluster through a critical hinge hub F592.

Local Frustration Patterns in the Ensembles of SARS-CoV-2 S-G614 States: Mutational Frustration Neutrality of Variant Sites and Modulation of Frustration in the Open States

We employed the conformational ensembles of the multiple conformational states of the S-G614 trimer to estimate the average local frustration profiles of the S residues and quantify the relationship between structural plasticity and degree of mutational frustration. This analysis is based on scanning conformational ensembles by local frustratometer^{147,148} which computes the local frustration index based on the contribution of a residue or residue pairs to the energy in a given conformation as compared to what it would contribute in decoy conformations. In this approach by performing mutations and changing local environments, the mean and variance of the energies of possible decoys relative to the native conformation are evaluated.¹⁴⁷⁻¹⁵¹ Previous studies by Wolynes and colleagues discovered that minimally frustrated residues are typically enriched in the rigid protein core, while a large fraction of residues are characterized by neutral frustration, i.e. these positions residue near the center of the distribution of possible energies in decoy states.¹⁴⁹⁻¹⁵¹ At the same time, it was found that residues near the binding sites the proteins are enriched in highly frustrated interactions.

We first examined the distribution of local mutational frustration in different substates of the S-G614 protein, particularly focusing on G614 position as well as sites targeted by common circulating variants in the RBD (K417, E484, N501) (Figure 6). The role of frustration in the RBD regions is particularly interesting in light of recent evidence that the disordered/highly flexible regions are important for mediating allostery and binding to multiple protein partners, where energetic frustration at the molecular interfaces may serve as a mechanism for allosteric regulation [96-99].

Figure 6. The distribution of local mutational frustration in the closed and open forms of the SARS-CoV-2 S-G614 variant. (A-C) The relative density of highly frustrated, neutrally frustrated and minimally frustrated residues respectively in the closed S-G614 conformations. The distributions were constructed by averaging computations over snapshots from simulations of the closed states (pdb ids 7KRQ and 7BNM). (D-F) The relative density of highly frustrated, neutrally frustrated and minimally frustrated residues respectively in the open S-G614 conformations. The distributions were constructed by averaging computations over snapshots from simulations of all 1 RBD-up and 2 RBD-up states (pdb ids 7KRR, 7BNN, 7BNO, 7EAZ, 7REB0, 7EB3, 7EB4, 7EB5).

First, we evaluated the cumulative distribution of local frustration index in the closed and open S-G614 states by averaging the residue-based local mutational frustration over dynamic ensembles of conformations (Figure 6). Interestingly, similar distributions of highly frustrated, neutral and minimally frustrated states respectively were seen in the closed (Figure 6A-C) and

open forms (Figure 6D-F). Notably, the population of highly frustrated conformations is relatively minor, but we observed a longer density tail of highly frustrated values for the distribution in the open states (Figure 6D). This may reflect a considerable conformational heterogeneity of the multiple open conformations where the frustrated sites correspond to highly exposed positions in the NTD and RBD. Importantly, we observed the dominant population of neutrally frustrated sites in both closed and open forms as well as a significant density of minimally frustrated regions. A relatively small fraction of highly frustrated positions as compared to neutral and minimally frustrated residues is an important observation of this analysis. This suggested that while the exposed RBD and NTD regions in the open states are fairly flexible, the degree of mutational frustration and adaptability in these positions may be moderate to enable structural stability and binding with the host receptor.

A comparison of local mutational frustration indexes for key functional positions including sites of variants (K417, E484, and N501), hinge position A570 and G614 revealed interesting shifts between the closed and open forms of S-G614 trimer (Figure 7). The high frustration index for these positions is uniformly small and similar in both closed and open conformations (Figure 7A,D). While this may be expected for hinge positions A570 and G614, it was somewhat revealing for the RBD positions, especially mobile E484 and N501 residues. It is of some notice that the relative high frustration density for G614 site becomes even smaller in the open form, indicating stabilizing effect of G614 residue in the receptor-accessible states (Figure 7A,D). The relative density of neutral frustration is pronounced for all sites, especially for N501 and G614 positions, showing no appreciable differences between the closed and open states (Figure 7B,E).

Figure 7. A comparison of the ensemble-averaged local mutational frustration between closed and open forms for sites of mutational variants and hinge positions (K417, E484, N501, A570, and G614) (A-C) The relative density of highly frustrated, neutrally frustrated and minimally frustrated residues respectively for the functional sites in the closed state (pdb id 7KRQ) and 1 RBD-up open state (pdb id 7KRR). (D-F) The relative density of highly frustrated, neutrally frustrated and minimally frustrated residues respectively for the functional sites in the closed state (pdb id 7BNM) and 2 RBD-up open state (pdb id 7BNO).

This suggested that sites of important circulating variants located in the flexible RBD regions showed a preponderance towards neutral mutational frustration. An appreciable density of minimal frustration for E484, A570 and especially G614 positions was also found (Figure 7C,F). In particular, we observed the increased minimal frustration of G614 in the ensemble of open conformations. Hence, the interactions of this key functional residue may become less frustrated in the open form as compared to the closed states. This is an interesting observation,

particularly given that the minimally frustrated interactions are often excluded from the vicinity of binding residues in globular proteins [94,95]. According to our results, the conformational plasticity of the RBD-up ensembles can allow for thermal rearrangements of the inter-protomer contact and repacking near G614 position, providing means for neutral-to-minimum frustration level in the mutational site. We argue that this allows for moderately suboptimal inter-protomer interactions leading to multiple open substates, each displaying specific frustration patterns, which can be exploited to control binding with the ACE2 and to multiple partners.

According to our data, the relative density of neutral frustration and minimal frustration can be similar for G614 position in the ensemble of open states. By modulating the frustration level and local dynamic environment around the A570/G614 hinge region, D614G mutation may modulate the thermodynamic equilibrium between the closed and open states as well as level of functional plasticity in the open substates required for allosteric function and/or binding.

Mutational Scanning of Protein Stability the SARS-CoV-2 S-614 Conformational States Reveals Energetic Effects of the D614G Mutation

We employed the equilibrium ensembles generated from CG-CABS simulations of the SARS-CoV-2 S protein structures to perform mutational scanning analysis of the S-G614 proteins (Figure 8). The questions addressed in this analysis are to examine the effect of the D614G mutation on different functional forms of the S-G614 protein and quantify the mutation-induced mechanism of S protein stabilization. The protein stability $\Delta\Delta G$ changes were computed by averaging the results of computations over 1,000 samples obtained from simulation trajectories. We constructed mutational sensitivity heatmaps for the major sites of circulating variants in the RBD (K417, E484, and N501) and G614 position in the S-G614

closed and open trimer forms. Mutational cartography revealed several interesting trends and highlighted the energetic preferences of G614 site (Figure 8). Interestingly, for the constrained S-G614 states, modifications in this position generally resulted in appreciable destabilization changes, while protein stability changes are only moderately perturbed by mutations in K417, E484 and N501 sites. Despite different level of structural plasticity in the examined closed and open states, mutational scanning map showed that the local structural environment around G614 is still fairly adaptable (Figure 8). Indeed, while modifications in the G614 position generally produced destabilizing free energy changes, the magnitude of these changes was relatively moderate. Notably, the magnitude of mutation-induced changes was smaller in the flexible open S-G614 states as compared to more constrained conformations (Figure 8). These findings are consistent with the mutational frustration analysis, indicating that the local environment near G614 is neutrally-to-minimally frustrated in the open conformations. In particular, the structural map showed a significant degree of energetic tolerance to G614 mutations in one of the 1 RBD-up open states (pdb id 7EB0) and 2 RBD-up conformations (Figure 8). At the same time, mutational scanning maps for the more dynamic S-G614 states displayed a pattern of moderate destabilization or neutral energetic changes for the RBD positions K417, E484 and N501 (Figure 8). The mutational cartography analysis revealed that a moderate level of energetic frustration and suboptimal interactions are characteristic not only of the sites of circulating variants located in the RBD but also for the D614G position. These energetic results are consistent and reciprocate the mutational frustration analysis, showing that the disordered and structured regions are enriched in neutrally (moderately) frustrated interactions. Our analysis suggested that D614G mutation may elicit sub-optimal interaction contacts and moderate level of frustration near the inter-protomer interfaces and hinge clusters.

Figure 8. Ensemble-based mutational profiling of the SARS-CoV-2 S-G614 stability. The mutational scanning heatmaps for the multiple SARS-CoV-2 S-G614 states at positions K417, E484, N501, and G614. The heatmaps show the computed binding free energy changes for 19 single mutations on the sites of variants. The squares on the heatmap are colored using a 3-colored scale – from pink to white and to green, with green indicating the largest unfavorable effect on binding.. The standard errors of the mean for binding free energy changes were based on 5 independent CG trajectories for each of the S-G614 states and different number of selected samples from a given trajectory (500, 1,000 and 2,000 samples) are ~ 0.15 - 0.22 kcal/mol using averages over different trajectories and ≤ 0.15 kcal/mol from computations based on different number of samples from a single trajectory.

We argue that the energetic frustration near G614 could allow for communication between hinge clusters and allosterically modulate the functional transitions of the S1 subunit. This landscape-based mechanism of mutation-induced energetic frustration in the S protein may

result in the greater adaptability and the emergence of multiple conformational substates in the open form. The results also suggested an interesting interplay between mutation-induced protein stability and local frustration patterns. It is particularly instructive to interpret these relationships in the context of allosteric regulation models centered on the energetic frustration that could emerge at the inter-domain interfaces.¹⁵²⁻¹⁵⁵ In our case, the D614G mutation may introduce the level of energetic frustration and conformational plasticity near the inter-protomer and inter-domain interfaces that would allow for efficient modulation of allosteric couplings between the hinge clusters without compromising the structural integrity of the S protein. Through this frustration-inspired allosteric mechanism, the D614G mutation can impose allosteric control over functional movements and conformational diversity of the RBD regions.

A more detailed analysis of mutational scanning in the constrained S-G614 states showed significant destabilization upon modifications in both closed and open states. Interestingly, however, hydrophobic modifications to F/W may produce marginally improved stability (Figure S3). The stabilization predicted by these mutations may be due to the hydrophobic effect and improved non-specific packing near the mutational site. Interestingly, the magnitude of destabilization changes was larger in the intermediate close state, particularly for substitutions that change the size and chemical nature of the amino acid such as G614D/E/K (Figure S3). Somewhat smaller but still significant destabilization changes upon G614 mutations are also detected in the 1 RBD-up open state. These results suggested that conformational plasticity is present even in the constrained S-G614 states which is consistent with the increased flexibility of the S protein and preferences for the open state induced by G614. These results also corroborated with the frustration analysis, suggesting a moderate degree of frustration in the

G614 position for open states. The range of destabilization free energy changes induced by mutations in the G614 position was considerably reduced in more flexible open structures determined by Benton and colleagues⁶³ (Figure S4). While the destabilization changes remain relatively pronounced in the closed form, the protein stability energy differences were very small in the 1 RBD-up and 2 RBD-up open states (Figure S4). The range of free energy $\Delta\Delta G$ changes where some modifications could even improve stability (G614D and G614N) was less than 1 kcal/mol, suggesting that these wide and mobile open conformations feature a considerable degree of conformational plasticity and local frustration near the mutational site. Accordingly, mutational tolerance of the S-G614 structures could facilitate transitions to the ensemble of open conformations with significant degree of conformational heterogeneity. We also detailed protein stability changes induced by G614 mutations in the ensemble of multiple conformational substates of the S-D614⁶⁵ (Figure S5). For these conformations, we generally observed moderate destabilization changes $\Delta\Delta G \sim 1.0\text{-}1.5$ kcal/mol. In particular, we noticed that modifications G614E/D/K and G614 substitutions to larger hydrophobic residues were highly unfavorable (Figure S5). Hence, these S-G614 open substates featured only moderate degree of local frustration and signs of minimal frustration as the local interactions near mutational site were more optimal (Figure S5). These results supported our hypothesis that conformational plasticity of the RBD-up structures can allow for neutral-to-minimum frustration level near the mutational G614 site. The suboptimal inter-protomer interactions in the multiple open substates featuring moderate frustration can allow for modulation of functional movements and may promote conformational transitions to the ensemble of open S-G614 states.

Dynamic Network Modeling and Ensemble-Based Community Analysis Identify Allosteric Interaction Networks in the SARS-CoV-2 Spike Mutants

Network-centric models of protein structure and dynamics used in the current study can also provide a complementary perspective to physics-based landscape models and allow for quantitative analysis of allosteric changes, identification of regulatory control centers and mapping of allosteric communication pathways. Using this framework, the residue interaction networks in the SARS-CoV-2 spike trimer structures were built using a graph-based representation of protein structures^{124,125} in which residue nodes are interconnected through both dynamic¹²⁶ and coevolutionary correlations.¹²⁷ By employing perturbation-based network modeling, we computed ensemble-averaged distributions of the betweenness centrality based on average shortest path length (ASPL) change upon removing/mutating individual nodes (Figure 9). Using these parameters, we identify mediating centers of allosteric interactions in the SARS-CoV-2 mutant structures. The residue interaction networks were also divided into local interaction communities in which residues are densely interconnected whereas residues from different communities may be weakly connected through the inter-modular links.

We first constructed the distribution of the short path betweenness profile for the structurally constrained S-G614 conformations (including closed, intermediate and partly open states) (Figure 9A). These S-G614 mutant structures that were solved with structurally ordered 630 loop (residues 617 to 644) and FPPR region (residues 823 to 862).⁶¹ We noticed that the network centrality profile displayed appreciable peaks in these regions (Figure 9A). These observations are consistent with structural studies indicating that the structured 630 loop in the S-G614 trimer can promote the tighter packing among three protomers and also stabilizes the couplings with CTD2 which inhibits the release of the N-terminal segment of S2 and blocking

S1 dissociation.⁶¹ The short betweenness profiles for the more rigid S-G614 states showed sharp and narrow peaks in the CTD1 region corresponding to residues D568 A570, T572, and F592 in the CTD1 region (Figure 9A). Several other peaks along the network betweenness profile corresponded to hinge centers F318, V539, L570, I572, T547, F592, S750, I856 and L962 residues (Figure 9A). Together, high centrality clusters could mediate allosteric couplings between the 630 loop, CTD1 residues and RBD regions. Notably, these regulatory points are located near the inter-protomer and inter-domain contact interfaces. The emergence of sharp single peaks in the distribution aligned with mediating centers in the S-G614 trimer signaled those allosteric communications in the structurally constrained S-G614 closed and open states may proceed through a fairly narrow pathway connecting key mediating sites located in CTD1 regions. In network terms, this would allow for efficient and direct signaling but could imply a degree of vulnerability to mutational perturbations of key mediating centers.

A noticeable change in the density of mediating residues was seen for the centrality distributions obtained for more flexible S-G614 open states (Figure 9B). For these structures, we observed the broadening of the density peaks in the CTD1 region and in the flexible 28-residue N2R linker spanning residues 306–334 that connects the NTD (residues 27–305) and RBD (residues 335–521) and stacks against the SD1 (residues 529–591) and SD2 (residues 592–697) regions. Significantly, the bifurcated distribution featured fairly broad peaks for the N2R and CTD1 regions, suggesting that the conformational flexibility of the S-G614 open states would translate to a much broader and arguably more robust ensemble of communication paths in the S-G614 structure. Notably, for all S-G614 open conformations, the G614 position was associated with a pronounced distribution peak. A similar pattern was obtained from the network analysis of the multiple substates of the open S-G614 conformations (Figure 9C).

Figure 9. Network centrality analysis of the SARS-CoV-2 S-G614 trimers. (A) The ensemble-averaged short path betweenness centrality for the structure of SARS-CoV-2 S-G614 in the closed state, pdb id 7KRQ (green lines), the intermediate form, pdb id 7KRS (red lines) and open form, pdb id 7KRR (blue lines). (B) The ensemble-averaged short path betweenness centrality for the structure of SARS-CoV-2 S-G614 in the closed state, pdb id 7BNM (green lines), the 1 RBD-up open form, pdb id 7BNN (red lines) and 2 RBD-up open form, pdb id 7BNO (blue lines). (B) The ensemble-averaged short path betweenness centrality for the structure of SARS-CoV-2 S-G614 in the 1 RBD-up open substate form, pdb id 7EB0 (green lines), 1 RBD-up substate form, pdb id 7EB3 (red lines), and 2 RBD-up open form, pdb id 7EB4 (blue lines). The SARS-CoV-2 S mutant trimers are shown in ribbons. The residue-based profiles are shown for protomers of the S trimer in green, red, and blue lines. The betweenness centrality profiles based on ASPL change upon removing individual nodes are shown for the structures.

In these structures, we also observed two very broad and distinct peaks of the short path centrality aligned with N2R and CTD1 regions, strongly arguing for a broad ensemble of signal transmission routes in the conformationally mobile open S-G614 states. The observed network signatures imply that structurally rigid hinge positions in CTD1 (A570, T572, F592 and G614) are coupled with more flexible residues to form a broad “tube” of allosteric communication routes. This would provide a significant resiliency and robustness to signaling in the S-G614 structures, making the allosteric function of the S protein robust against perturbations and mutations. The results also suggested an important role to the N2R linker that connects the NTD regions to the RBD residues within a given protomer. Rather than just being a connector, this 28-residue linker could also function as a modulator of allosteric conformational changes that are critical for the ACE2 receptor engagement. The important conclusion of this analysis is that by inducing conformational heterogeneity of the open S form, the D614G variant allows for robust allosteric communications between the site of mutation, CTD1, NTD and RBD regions. In this scenario, allosteric signal initiated near the mutational site in the open state, would propagate via a broad “tube” of communication routes to the RBD regions, making allosteric interactions robust to perturbations and variations in the ensemble of the S conformations.

We also performed community analysis to monitor dynamic changes in the modular organization of the inter-protomer interfaces. The community-based analysis and modularity assessment of allosteric interaction networks is based on the notion that groups of residues that form local interacting communities are expected to be highly correlated and can switch their conformational states cooperatively. As a result, long-range allosteric interaction signals can be transmitted through a hierarchical chain of local communities on the dynamic interaction networks.

Figure 8. Network community analysis of the SARS-CoV-2 S-G614 closed and open states. Structural mapping of key local communities (residues comprising these modules are in red spheres) in the structure of SARS-CoV-2 S-G614 in the closed state, pdb id 7KRQ (A), and the structure 1 RBD-up open form, pdb id 7BNN (C). The SARS-CoV-2 S mutant protomer is shown in ribbons depicting the “down protomer” for the closed state on panel (A) and “up protomer” for the partially open state on panel (C). The S-G614 structures are shown in ribbons with the chain A in green, chain B in cyan, and chain C in blue colors. (B) A close-up of the inter-protomer hinge communities formed in the SARS-CoV-2 S-G614 in the closed state, pdb id 7KRQ. (D) A close-up of the inter-protomer hinge communities formed in the SARS-CoV-2 S-G614 in the 1 RBD-up open state, pdb id 7BNN. The communities are shown in spheres colored according to the respective chain. The residues that comprise the local communities are annotated. Note a partial fragmentation of local communities in the open form.

The community analysis of the S-G614 structures is used as proxy for measuring protein stability and communication efficiency in performing allosteric functions. This analysis provides also a global metric to measure changes in connectivity and interaction between subdomains and inter-protomer association in the SARS-CoV-2 S structures. In the closed S-G614 rigid trimer, the interfacial communities in the hinge region include A614-C855-C589-C592; A568-A574-C854; A614-A740-A857-B592; B775-B864-A665; A592-C737-C855 (Figure 10A,B). The hinge positions in the closed state anchor the inter-protomer communities A568-A574-C854; B855-C589-C592; A614-A592-A589-B835-B854; A614-B740-B737-A592; and A592-C737-C855 (Figure 10A,B) and control propagation of allosteric signal between S1 and S2 subdomains through a fairly narrow channel formed by these communities in the CTD1 region. In the S-G614 trimer, G614 forms community with F592 and P589 of the same protomer, K835 and K854 of the adjacent protomer (Figure 10A,B). This rewiring allows for strong local coupling of G614 and hinge center F592 and provides a “stapling” lock for the rigid close conformation. This analysis provided an additional support to the notion that D614G can function as an important allosteric hub for mediating the increased flexibility in the SD1-S2 hinge regions. An emerging concept central to computational models of allostery is identification of conformational switch centers that mediate long-range communications and allosteric pathways.¹⁵⁶⁻¹⁵⁸ Our analysis suggested that these inter-protomer communities near hinge centers could function as such allosteric security switches that facilitate large scale transitions to the open form. We observed several common changes in the inter-protomer communities of the more flexible 1 RBD-up conformations (Figure 10C,D). First, G614 and P589 weaken their interactions with the communities A614-B855-C589-C592 and G614 becomes further decoupled from the community A614-B740-B737-A592 (Figure 10C,D).

Second, we found that the local intra-protomer community formed by residues A570, D568 and D574 is tightly coupled to the network of the interfacial communities in the closed S-G614 trimer (Figure 10A,B), but becomes separated from the inter-protomer communities in the 1 RBD-up open form due to greater conformational variability in this region (Figure 10C,D). In the ensemble-averaged analysis, the weakened contacts between these communities become further exacerbated which leads to fragmentation of the major hinge cluster formed by F318, F592, A570, I572, G614 and Y855 residues. As a result, in the open S-G614 states, G614 site is community-coupled to F855 and M740 positions of the same protomer and hinge site F592 of the adjacent protomer. In addition, an important hinge center A570 becomes disengaged from the inter-protomer hinge cluster that may release some constraints on movements in the S1 regions and promote greater conformational flexibility of the open protomers. These observations from the community analysis are consistent with the local frustration profiling showing modulation of the frustration level and local dynamic environment around the A570/G614 hinge region, which may increase functional plasticity in the open substates required for allosteric function and/or binding. Structural mapping of communities across different states indicated that F592 could anchor the central inter-protomer communities and link them to G614 site. For more flexible 1 RBD-up conformational states of S-G614, we observed loosening of these communities around F592, suggesting that this position could effectively alter communications in the S-G614 trimer structures (Figure 10). At the same time, we observed some reorganization of the interfacial communities in the SD1-S2 region that harbors most of the hinge positions (Figure 10C,D). These network changes reflect strengthening of the interactions in the S2 subunit, while considerable mobility of the S1 regions can induce a reorganization of local communities. The results indicated that the regulatory control points could link together

different local communities to form allosteric communication paths in the system. Hence, effective allosteric communications in functional states of the S-G614 trimer can be mediated by clusters of high centrality mediating residues. The network analysis revealed that the high centrality centers could be surrounded by clusters of neighboring residues featuring also fairly high betweenness values. As a result, functional sites critical for signal transmission between site of mutation and RBD regions could be shielded by neighboring residues that have sufficient communication capabilities to ensure broad communication routes and resilience in the fluctuating protein environment. Although mutations of functional residues may often result in a loss of activity, some of these changes could be rescued by neighboring residues that assume additional functional responsibilities in the altered interaction network.¹⁵⁶

Conclusions

The results of this study demonstrated that the D614G mutation may induce a considerable conformational adaptability of the open states without compromising folding stability and integrity of the spike protein. Using the local frustration profiling of the S-G614 conformational ensembles we found that conformational plasticity of the open S conformations can allow for moderate thermal rearrangements of the inter-protomer interactions and favor neutral frustration level near the mutational site and sites of circulating variants. We proposed a mechanism in which suboptimal inter-protomer interactions displaying specific frustration patterns may be exploited to control binding with the ACE2 and multiple binding partners. Functional dynamics analysis suggested that G614 site may serve as a versatile regulatory switch point that links or decouples the network of hinge clusters. The ensemble-based mutational energetic scanning of protein stability for the S-G614 substates suggests a significant mutational plasticity of the open S-G614 trimer. This study suggests that modulation

of the energetic frustration at the inter-protomer interfaces by D614G mutation can serve as a mechanism for allosteric regulation in which the dynamic couplings between the site of mutation and the inter-protomer hinge of functional motions would modulate the inter-domain interactions, global changes in mobility and the increased stability of the open form. The network analysis showed that by promoting the enhanced mobility of the open states the D614G variant can induce allosteric communication routes that are robust to perturbations in the dynamic protein environment. The results of this study suggest that conformational plasticity and frustration-driven allostery of the S-614 protein are important functional mechanisms of this highly transmissible spike variant.

DATA AND SOFTWARE AVAILABILITY

Crystal structures were obtained and downloaded from the Protein Data Bank (<http://www.rcsb.org>), accession numbers 7KRQ, 7BNM (the cryo-EM structures of SARS-CoV-2 S-G614 in the closed state), 7KRQ (the S-G614 intermediate state), 7KRR, 7BNN, 7EAZ, 7EB0, 7EB3 (1 RBD-up open states), and 7BNO, 7EB4, 7EB5 (2 RBD-up open states). All simulations were performed using NAMD 2.13 package that was obtained from website <https://www.ks.uiuc.edu/Development/Download/>. All simulations were performed using the all-atom additive CHARMM36 protein force field that can be obtained from http://mackerell.umaryland.edu/charmm_ff.shtml. The residue interaction network files in xml format were obtained for all structures using the Residue Interaction Network Generator (RING) program RING v2.0.1 freely available at <http://old.protein.bio.unipd.it/ring/>. The computations of network parameters were conducted using NAPS program available at <https://bioinf.iit.ac.in/NAPS/index.php> and Cytoscape 3.8.2 environment available that can be

downloaded from <https://cytoscape.org/download.html>. The rendering of SARS-CoV-2 S structures was done using the interactive visualization program UCSF ChimeraX package (<https://www.rbvi.ucsf.edu/chimerax/>) and Pymol (<https://pymol.org/2/>). The scripts and programs for computational analysis are available in the GitHub site <https://github.com/stevenagajanian/KInGAN>.

SUPPORTING INFORMATION

Figure S1 illustrates the domain organization for the full-length SARS-CoV-2 S protein. Figure S2 describes Functional dynamics of the SARS-CoV-2 S-D614 and S-G614 trimer structures in the locked closed, intermediate and open forms obtained using PCA of atomistic trajectories. Figure S3 shows the mutational sensitivity analysis of the G614 residue in the SARS-CoV-2 S-G614 closed state, intermediate state, and 1 RBD-up open form. Figure S4 details the mutational sensitivity analysis of the G614 residue in the SARS-CoV-2 S-G614 closed state, the 1 RBD-up state and 2 RBD-up open form. Figure S5 presents specifics of the mutational sensitivity analysis of the G614 residue in the ensemble of SARS-CoV-2 S-G614 open sub states.

This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

* Corresponding Author

Phone: 714-516-4586; Fax: 714-532-6048; E-mail: verkhivk@chapman.edu

The authors declare no competing financial interest.

Funding: This research received no external funding

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgment The author acknowledges support by the Kay Family Foundation Grant A20-0032.

ABBREVIATIONS

SARS, Severe Acute Respiratory Syndrome; RBD, Receptor Binding Domain; ACE2, Angiotensin-Converting Enzyme 2 (ACE2); NTD, N-terminal domain; RBD, receptor-binding domain; CTD1, C-terminal domain 1; CTD2, C-terminal domain 2; FP, fusion peptide; FPPR, fusion peptide proximal region; HR1, heptad repeat 1; CH, central helix region; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane anchor; CT, cytoplasmic tail.

References

- (1) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y., et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181*, 894-904.e9.
- (2) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 11727-11734.4
- (3) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veessler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281-292.e6.
- (4) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260-1263.
- (5) Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M., Jr.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020**, *369*, 1586-1592.
- (6) Hsieh, C. L.; Goldsmith, J. A.; Schaub, J. M.; DiVenere, A. M.; Kuo, H. C.; Javanmardi, K.; Le, K. C.; Wrapp, D.; Lee, A. G.; Liu, Y., et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **2020**, *369*, 1501-1505.
- (7) Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S. M. C.; Kopp, M.; Li, D.; Parks, R.; Hsu, A. L., et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* **2020**, *27*, 925-933.
- (8) McCallum, M.; Walls, A. C.; Bowen, J. E.; Corti, D.; Veessler, D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat. Struct. Mol. Biol.* **2020**, *27*, 942-949.

- (9) Xiong, X.; Qu, K.; Ciazynska, K. A.; Hosmillo, M.; Carter, A. P.; Ebrahimi, S.; Ke, Z.; Scheres, S. H. W.; Bergamaschi, L.; Grice, G. L., et al. A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat. Struct. Mol. Biol.* **2020**, *27*, 934-941.
- (10) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588*, 327-330.
- (11) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S.; Blanc, F. E. C.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C., et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, *370*, 203-208.
- (12) Lu, M.; Uchil, P. D.; Li, W.; Zheng, D.; Terry, D. S.; Gorman, J.; Shi, W.; Zhang, B.; Zhou, T.; Ding, S., et al. Real-time conformational dynamics of SARS-CoV-2 spikes on virus particles. *Cell Host Microbe*. **2020**, *28*, 880-891.e8.
- (13) Gavor, E.; Choong, Y. K.; Er, S. Y.; Sivaraman, H.; Sivaraman, J. Structural Basis of SARS-CoV-2 and SARS-CoV Antibody Interactions. *Trends Immunol* **2020**, *41*, 1006-1022.
- (14) Finkelstein, M. T.; Mermelstein, A. G.; Parker Miller, E.; Seth, P. C.; Stancofski, E. D.; Fera, D., Structural Analysis of Neutralizing Epitopes of the SARS-CoV-2 Spike to Guide Therapy and Vaccine Design Strategies. *Viruses* **2021**, *13*, 134.
- (15) Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E., et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **2020**, *588*, 682-687.
- (16) Barnes, C. O.; West, A. P., Jr.; Huey-Tubman, K. E.; Hoffmann, M. A. G.; Sharaf, N. G.; Hoffman, P. R.; Koranda, N.; Gristick, H. B.; Gaebler, C.; Muecksch, F., et al. Structures of

Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **2020**, *182*, 828-842.e16.

(17) Hansen, J.; Baum, A.; Pascal, K. E.; Russo, V.; Giordano, S.; Wloga, E.; Fulton, B. O.; Yan, Y.; Koon, K.; Patel, K., et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **2020**, *369*, 1010-1014.

(18) Wu, Y.; Wang, F.; Shen, C.; Peng, W.; Li, D.; Zhao, C.; Li, Z.; Li, S.; Bi, Y.; Yang, Y., 2et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **2020**, *368*, 1274-1278.

(19) Shi, R.; Shan, C.; Duan, X.; Chen, Z.; Liu, P.; Song, J.; Song, T.; Bi, X.; Han, C.; Wu, L., et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **2020**, *584*, 120-124.

(20) Ju, B.; Zhang, Q.; Ge, J.; Wang, R.; Sun, J.; Ge, X.; Yu, J.; Shan, S.; Zhou, B.; Song, S., et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **2020**, *584*, 115-119.

(21) Du, S.; Cao, Y.; Zhu, Q.; Yu, P.; Qi, F.; Wang, G.; Du, X.; Bao, L.; Deng, W.; Zhu, H., et al. Structurally Resolved SARS-CoV-2 Antibody Shows High Efficacy in Severely Infected Hamsters and Provides a Potent Cocktail Pairing Strategy. *Cell* **2020**, *183*, 1013-1023.e13.

(22) Baum, A.; Fulton, B. O.; Wloga, E.; Copin, R.; Pascal, K. E.; Russo, V.; Giordano, S.; Lanza, K.; Negron, N.; Ni, M., et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **2020**, *369*, 1014-1018.

(23) Yu, F.; Xiang, R.; Deng, X.; Wang, L.; Yu, Z.; Tian, S.; Liang, R.; Li, Y.; Ying, T.; Jiang, S. Receptor-binding domain-specific human neutralizing monoclonal antibodies against SARS-CoV and SARS-CoV-2. *Signal Transduct Target Ther.* **2020**, *5*, 212.

- (24) Piccoli, L.; Park, Y. J.; Tortorici, M. A.; Czudnochowski, N.; Walls, A. C.; Beltramello, M.; Silacci-Fregni, C.; Pinto, D.; Rosen, L. E.; Bowen, J. E., et al. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **2020**, *183*, 1024-1042.e21.
- (25) Ge, J.; Wang, R.; Ju, B.; Zhang, Q.; Sun, J.; Chen, P.; Zhang, S.; Tian, Y.; Shan, S.; Cheng, L., et al. Antibody neutralization of SARS-CoV-2 through ACE2 receptor mimicry. *Nat Commun* **2021**, *12*, 250.
- (26) Ku, Z.; Xie, X.; Davidson, E.; Ye, X.; Su, H.; Menachery, V. D.; Li, Y.; Yuan, Z.; Zhang, X.; Muruato, A. E., et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat Commun* **2021**, *12*, 469.
- (27) Yuan, M.; Liu, H.; Wu, N. C.; Lee, C. D.; Zhu, X.; Zhao, F.; Huang, D.; Yu, W.; Hua, Y.; Tien, H., et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science* **2020**, *369*, 1119-1123.
- (28) Zhou, D.; Duyvesteyn, H. M. E.; Chen, C. P.; Huang, C. G.; Chen, T. H.; Shih, S. R.; Lin, Y. C.; Cheng, C. Y.; Cheng, S. H.; Huang, Y. C., et al. Structural basis for the neutralization of SARS-CoV-2 by an antibody from a convalescent patient. *Nat Struct Mol Biol* **2020**, *27*, 950-958.
- (29) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y., et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **2020**, *369*, 650-655.
- (30) Brouwer, P. J. M.; Caniels, T. G.; van der Straten, K.; Snitselaar, J. L.; Aldon, Y.; Bangaru, S.; Torres, J. L.; Okba, N. M. A.; Claireaux, M.; Kerster, G., et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **2020**, *369*, 643-650.

- (31) Lv, Z.; Deng, Y. Q.; Ye, Q.; Cao, L.; Sun, C. Y.; Fan, C.; Huang, W.; Sun, S.; Sun, Y.; Zhu, L, et al. Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody. *Science* **2020**, *369*, 1505-1509.
- (32) Pinto, D.; Park, Y. J.; Beltramello, M.; Walls, A. C.; Tortorici, M. A.; Bianchi, S.; Jaconi, S.; Culap, K.; Zatta, F.; De Marco, A., et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **2020**, *583*, 290-295.
- (33) Tortorici, M. A.; Beltramello, M.; Lempp, F. A.; Pinto, D.; Dang, H. V.; Rosen, L. E.; McCallum, M.; Bowen, J.; Minola, A.; Jaconi, S., et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* **2020**, *370*, 950-957.
- (34) Asarnow, D.; Wang, B.; Lee, W. H.; Hu, Y.; Huang, C. W.; Faust, B.; Ng, P. M. L.; Ngoh, E. Z. X.; Bohn, M.; Bulkley, D., et al. Structural insight into SARS-CoV-2 neutralizing antibodies and modulation of syncytia. *Cell* **2021**, *184*, 3192-3204.e16.
- (35) Hurlburt, N. K.; Seydoux, E.; Wan, Y. H.; Edara, V. V.; Stuart, A. B.; Feng, J.; Suthar, M. S.; McGuire, A. T.; Stamatatos, L.; Pancera, M., Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation. *Nat Commun* **2020**, *11*, 5413.
- (36) Gaebler, C.; Wang, Z.; Lorenzi, J. C. C.; Muecksch, F.; Finkin, S.; Tokuyama, M.; Cho, A.; Jankovic, M.; Schaefer-Babajew, D.; Oliveira, T. Y., et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **2021**, *591*, 639-644.
- (37) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H. D.; Diggins, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C., et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182*, 1295-1310.e20.

- (38) Greaney, A. J.; Starr, T. N.; Gilchuk, P.; Zost, S. J.; Binshtein, E.; Loes, A. N.; Hilton, S. K.; Huddleston, J.; Eguia, R.; Crawford, K. H. D., et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **2021**, 29, 44-57.e9.
- (39) Greaney, A.J.; Loes, A.N.; Crawford, K.H.D.; Starr, T.N.; Malone, K.D.; Chu, H.Y.; Bloom, J.D. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies *Cell Host Microbe* **2021**, 29, 463-476.e6.
- (40) Starr, T. N.; Greaney, A. J.; Addetia, A.; Hannon, W. W.; Choudhary, M. C.; Dingens, A. S.; Li, J. Z.; Bloom, J. D. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **2021**, 371,850-854.
- (41) Greaney, A. J.; Starr, T. N.; Barnes, C. O.; Weisblum, Y.; Schmidt, F.; Caskey, M.; Gaebler, C.; Cho, A.; Agudelo, M.; Finkin, S., et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat Commun* **2021**, 12, 4196.
- (42) Starr, T. N.; Czudnochowski, N.; Liu, Z.; Zatta, F.; Park, Y. J.; Addetia, A.; Pinto, D.; Beltramello, M.; Hernandez, P.; Greaney, A. J., et al. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **2021**, 597, 97-102.
- (43) Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C. O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J. A., et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **2021** 592, 616-622.
- (44) De Gasparo, R.; Pedotti, M.; Simonelli, L.; Nickl, P.; Muecksch, F.; Cassaniti, I.; Percivalle, E.; Lorenzi, J. C. C.; Mazzola, F.; Magrì, D., et al. Bispecific IgG neutralizes SARS-CoV-2 variants and prevents escape in mice. *Nature* **2021**, 593 (7859), 424-428.

- (45) Cai, Y.; Zhang, J.; Xiao, T.; Lavine, C. L.; Rawson, S.; Peng, H.; Zhu, H.; Anand, K.; Tong, P.; Gautam, A., et al. Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* **2021**, 373, 642-648.
- (46) McCallum, M.; Bassi, J.; De Marco, A.; Chen, A.; Walls, A. C.; Di Iulio, J.; Tortorici, M. A.; Navarro, M. J.; Silacci-Fregni, C.; Saliba, C., et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* **2021**, 373, 648-654.
- (47) Gobeil, S. M.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M. F.; Manne, K.; Li, D.; Wiehe, K., et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* **2021**, 373, eabi6226.
- (48) Yuan, M.; Huang, D.; Lee, C. D.; Wu, N. C.; Jackson, A. M.; Zhu, X.; Liu, H.; Peng, L.; van Gils, M. J.; Sanders, R. W., et al. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **2021**, 373, 818-823.
- (49) Wang, L.; Zhou, T.; Zhang, Y.; Yang, E. S.; Schramm, C. A.; Shi, W.; Pegu, A.; Oloniniyi, O. K.; Henry, A. R.; Darko, S., et al. Ultrapotent antibodies against diverse and highly transmissible SARS-CoV-2 variants. *Science* **2021**, 373, eabh1766. 2
- (50) Tortorici, M. A.; Czudnochowski, N.; Starr, T. N.; Marzi, R.; Walls, A. C.; Zatta, F.; Bowen, J. E.; Jaconi, S.; Di Iulio, J.; Wang, Z., et al. Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature* **2021**, 597, 103-108.
- (51) Dong, J.; Zost, S. J.; Greaney, A. J.; Starr, T. N.; Dingens, A. S.; Chen, E. C.; Chen, R. E.; Case, J. B.; Sutton, R. E.; Gilchuk, P., et al. Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. *Nat Microbiol* **2021**, 6, 1233-1244.
- (52) Cho, H.; Gonzales-Wartz, K. K.; Huang, D.; Yuan, M.; Peterson, M.; Liang, J.; Beutler, N.; Torres, J. L.; Cong, Y.; Postnikova, E., et al. Bispecific antibodies targeting distinct regions of

the spike protein potentially neutralize SARS-CoV-2 variants of concern. *Sci Transl Med* **2021**, 13, eabj5413.

(53) Li, T.; Xue, W.; Zheng, Q.; Song, S.; Yang, C.; Xiong, H.; Zhang, S.; Hong, M.; Zhang, Y.; Yu, H., et al. Cross-neutralizing antibodies bind a SARS-CoV-2 cryptic site and resist circulating variants. *Nat Commun* **2021**, 12, 5652.

(54) Plante, J. A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B. A.; Lokugamage, K. G.; Zhang, X.; Muruato, A. E.; Zou, J.; Fontes-Garfias, C. R., et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2020**, 592, 116-121.

(55) Hou, Y. J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinno, K. H., 3rd; Leist, S. R.; Schäfer, A.; Nakajima, N.; Takahashi, K., et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **2020**, 370, 1464-1468.

(56) Yurkovetskiy, L.; Wang, X.; Pascal, K. E.; Tomkins-Tinch, C.; Nyalile, T. P.; Wang, Y.; Baum, A.; Diehl, W. E.; Dauphin, A.; Carbone, C., et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* **2020**, 183, 739-751.e8.

(57) Gobeil, S. M.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Manne, K.; Stalls, V.; Kopp, M. F.; Henderson, R.; Edwards, R. J., et al. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep* **2021**, 34, 108630.

(58) Weissman, D.; Alameh, M. G.; de Silva, T.; Collini, P.; Hornsby, H.; Brown, R.; LaBranche, C. C.; Edwards, R. J.; Sutherland, L.; Santra, S., et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe* **2021**, 29, 23-31.

- (59) Zhang, L.; Jackson, C. B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B. D.; Rangarajan, E. S.; Pan, A.; Vanderheiden, A.; Suthar, M. S., et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* **2020**, *11*, 6013.
- (60) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B., et al. Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812-827.e19.
- (61) Zhang, J.; Cai, Y.; Xiao, T.; Lu, J.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rits-Volloch, S.; Sliz, P.; Chen, B. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **2021**, *372*, 525-530.
- (62) Juraszek, J.; Rutten, L.; Blokland, S.; Bouchier, P.; Voorzaat, R.; Ritschel, T.; Bakkers, M. J. G.; Renault, L. L. R.; Langedijk, J. P. M. Stabilizing the closed SARS-CoV-2 spike trimer. *Nat Commun* **2021**, *12*, 244.
- (63) Benton, D. J.; Wrobel, A. G.; Roustan, C.; Borg, A.; Xu, P.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc Natl Acad Sci U S A* **2021**, *118*, e2022586118.
- (64) Yan, R.; Zhang, Y.; Li, Y.; Ye, F.; Guo, Y.; Xia, L.; Zhong, X.; Chi, X.; Zhou, Q. Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell Res* **2021**, *31*, 717-719.
- (65) Yang, T. J.; Yu, P. Y.; Chang, Y. C.; Hsu, S. D. D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. *J Biol Chem* **2021**, *297*, 101238.

- (66) Huang, S. Y.; Kung, Y. A.; Huang, P. N.; Chang, S. Y.; Gong, Y. N.; Han, Y. J.; Chiang, H. J.; Liu, K. T.; Lee, K. M.; Chang, C. Y.; et al. Stability of SARS-CoV-2 Spike G614 Variant Surpasses That of the D614 Variant after Cold Storage. *mSphere* **2021**, *6*, e00104-e00121.
- (67) Ozono, S.; Zhang, Y.; Ode, H.; Sano, K.; Tan, T. S.; Imai, K.; Miyoshi, K.; Kishigami, S.; Ueno, T.; Iwatani, Y.; et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun* **2021**, *12*, 848.
- (68) Woo, H.; Park, S. J.; Choi, Y. K.; Park, T.; Tanveer, M.; Cao, Y.; Kern, N. R.; Lee, J.; Yeom, M. S.; Croll, T. I., et al. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B* **2020**, *124*, 7128-7137.
- (69) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S. et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722-1734.
- (70) Sztain, T.; Ahn, S.H.; Bogetti, A.T.; Casalino, L.; Goldsmith, J.A.; Seitz, E.; McCool, R.S.; Kearns, F.L.; Acosta-Reyes, F.; Maji, S., et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* **2021** *13*, 963-968.
- (71) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys J* **2021**, *120*, 1097-1104.
- (72) Sikora, M.; von Bülow, S.; Blanc, F. E. C.; Gecht, M.; Covino, R.; Hummer, G., Computational epitope map of SARS-CoV-2 spike protein. *PLoS Comput. Biol.* **2021**, *17*, e1008790.
- (73) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P., SARS-CoV-2 simulations

go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat Chem* **2021**, *13*, 651-659.

(74) Fatihi, S.; Rathore, S.; Pathak, A.K.; Gahlot, D.; Mukerji, M.; Jatana, N.; Thukral, L. A rigorous framework for detecting SARS-CoV-2 spike protein mutational ensemble from genomic and structural features. *Curr Res Struct Biol.* **2021**, *3*, 290-300.

(75) Ray, D.; Le, L.; Andricioaei, I. Distant residues modulate conformational opening in SARS-CoV-2 spike protein. *Proc Natl Acad Sci U S A* **2021**, *118*, e2100943118.

(76) Fallon, L.; Belfon, K. A. A.; Raguette, L.; Wang, Y.; Stepanenko, D.; Cuomo, A.; Guerra, J.; Budhan, S.; Varghese, S.; Corbo, C. P.; et al. Free Energy Landscapes from SARS-CoV-2 Spike Glycoprotein Simulations Suggest that RBD Opening Can Be Modulated via Interactions in an Allosteric Pocket. *J Am Chem Soc* **2021**, *143*, 11349-11360.

(77) Mansbach, R. A.; Chakraborty, S.; Nguyen, K.; Montefiori, D. C.; Korber, B.; Gnanakaran, S. The SARS-CoV-2 Spike variant D614G favors an open conformational state. *Sci Adv* **2021**, *7*, eabf3671.

(78) Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., Wang, Y., Hong, Q., Wang, S., Zhao, Q., et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* **2021**, *7*, eabe5575.

(79) Teruel, N.; Mailhot, O.; Najmanovich, R. J. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *PLoS Comput Biol* **2021**, *17*, e1009286.

(80) Yazhini, A.; Sidhanta, D. S. P.; Srinivasan, N. D614G substitution at the hinge region enhances the stability of trimeric SARS-CoV-2 spike protein. *Bioinformation* **2021**, *17*, 439-445.

- (81) Fernández, A., Structural Impact of Mutation D614G in SARS-CoV-2 Spike Protein: Enhanced Infectivity and Therapeutic Opportunity. *ACS Med Chem Lett* **2020**, *11*, 1667-1670.
- (82) Verkhivker, G.M. Molecular simulations and network modeling reveal an allosteric signaling in the SARS-CoV-2 spike proteins. *J. Proteome Res.* **2020**, *19*, 4587-4608.
- (83) Verkhivker, G. M.; Di Paola, L., Dynamic Network Modeling of Allosteric Interactions and Communication Pathways in the SARS-CoV-2 Spike Trimer Mutants: Differential Modulation of Conformational Landscapes and Signal Transmission via Cascades of Regulatory Switches. *J Phys. Chem. B.* **2021**, *125*, 850-873.
- (84) Verkhivker, G.M.; Di Paola, L. Integrated Biophysical Modeling of the SARS-CoV-2 Spike Protein Binding and Allosteric Interactions with Antibodies. *J. Phys. Chem. B.* **2021**, *125*, 4596-4619
- (85) Verkhivker, G.M.; Agajanian, S.; Oztas, D.Y.; Gupta, G. Comparative Perturbation-Based Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. *Biochemistry* **2021**, *60*, 1459-1484.
- (86) Verkhivker, G.; Agajanian, S.; Oztas, D.; Gupta, G. Dynamic Profiling of Binding and Allosteric Propensities of the SARS-CoV-2 Spike Protein with Different Classes of Antibodies: Mutational and Perturbation-Based Scanning Reveals the Allosteric Duality of Functionally Adaptable Hotspots. *J. Chem. Theory Comput.* **2021**, *17*, 4578-4598.
- (87) Verkhivker, G.M.; Agajanian, S.; Oztas, D.Y.; Gupta, G. Landscape-Based Mutational Sensitivity Cartography and Network Community Analysis of the SARS-CoV-2 Spike Protein Structures: Quantifying Functional Effects of the Circulating D614G Variant. *ACS Omega* **2021**, *6*, 16216-16233.

- (88) Verkhivker, G.M.; Agajanian, S.; Oztas, D.Y.; Gupta, G. Allosteric Control of Structural Mimicry and Mutational Escape in the SARS-CoV-2 Spike Protein Complexes with the ACE2 Decoys and Miniprotein Inhibitors: A Network-Based Approach for Mutational Profiling of Binding and Signaling. *J Chem Inf Model* **2021** 61, 5172-5191.
- (89) Verkhivker, G.M.; Agajanian, S.; Oztas, D.Y.; Gupta, G. Atomistic Simulations and In Silico Mutational Profiling of Protein Stability and Binding in the SARS-CoV-2 Spike Protein Complexes with Nanobodies: Molecular Determinants of Mutational Escape Mechanisms. *ACS Omega* **2021** 6 , 26354-26371.
- (90) Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235-242.
- (91) Rose, P. W.; Prlic, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Costanzo, L. D.; Duarte, J. M.; Dutta, S.; Feng, Z., et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, 45, D271-D281.
- (92) Hooft, R. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins* **1996**, 26, 363-376.
- (93) Hekkelman, M. L.; Te Beek, T. A.; Pettifer, S. R.; Thorne, D.; Attwood, T. K.; Vriend, G. WIWS: A protein structure bioinformatics web service collection. *Nucleic Acids Res.* **2010**, 38, W719-W723.
- (94) Fiser, A.; Sali, A. ModLoop: Automated modeling of loops in protein structures. *Bioinformatics* **2003**, 19, 2500-2501.
- (95) Fernandez-Fuentes, N.; Zhai, J.; Fiser, A. ArchPRED: A template based loop structure prediction server. *Nucleic Acids Res.* **2006**, 34, W173-W176.

- (96) Ko, J.; Lee, D.; Park, H.; Coutsiar, E. A.; Lee, J.; Seok, C. The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res.* **2011**, *39*, W210-W214.
- (97) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778-795.
- (98) Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460-1465.
- (99) Lombardi, L. E.; Marti, M. A.; Capece, L. CG2AA: backmapping protein coarse-grained structures. *Bioinformatics* **2016**, *32*, 1235-1237.
- (100) Bhattacharya, D.; Nowotny, J.; Cao, R.; Cheng, J. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res.* **2016**, *44*, W406-W409.
- (101) Watanabe, Y.; Berendsen, Z. T.; Raghvani, J.; Seabright, G. E.; Allen, J. D.; Pybus, O. G.; McLellan, J. S.; Wilson, I. A.; Bowden, T. A.; Ward, A. B.; Crispin, M. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **2020**, *11*, 2688.
- (102) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369*, 330-333.
- (103) Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* **2004**, *51*, 349-371.
- (104) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898-7936.
- (105) Kmiecik, S.; Kouza, M.; Badaczewska-Dawid, A.E.; Kloczkowski, A.; Kolinski, A. Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models. *Int. J. Mol. Sci.* **2018**, *19*, e3496.

- (106) Ciemny, M.P.; Badaczewska-Dawid, A.E.; Pikuzinska, M.; Kolinski, A.; Kmiecik, S. Modeling of disordered protein structures using monte carlo simulations and knowledge-based statistical force fields. *Int. J. Mol. Sci.* **2019**, *20*, e606.
- (107) Kurcinski, M.; Oleniecki, T.; Ciemny, M.P.; Kuriata, A.; Kolinski, A.; Kmiecik, S. CABS-flex standalone: A simulation environment for fast modeling of protein flexibility. *Bioinformatics* **2019**, *35*, 694-695.
- (108) Jamroz, M.; Orozco, M.; Kolinski, A.; Kmiecik, S. Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. *J. Chem. Theory Comput.* **2013**, *9*, 119-125.
- (109) Badaczewska-Dawid, A. E.; Kolinski, A.; Kmiecik, S. Protocols for fast simulations of protein structure flexibility using CABS-Flex and SURPASS. *Methods Mol. Biol.* **2020**, *2165*, 337-353.
- (110) Koukos, P.I.; Glykos, N.M. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J. Comput. Chem.* **2013**, *34*, 2310-2312.
- (111) Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E. Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* **2010**, *39*, 23-42.
- (112) Yang, L. W.; Rader, A. J.; Liu, X.; Jursa, C. J.; Chen, S. C.; Karimi, H. A.; Bahar, I. oGNM: online computation of structural dynamics using the gaussian network model. *Nucleic Acids Res.* **2006**, *34*, W24-W31.
- (113) Li, H.; Chang, Y. Y.; Lee, J. Y.; Bahar, I.; Yang, L. W. DynOmics: dynamics of structural proteome and beyond. *Nucleic Acids Res.* **2017**, *45*, W374-W380.
- (114) Eyal, E.; Lum, G.; Bahar, I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31*, 1487-1489.

- (115) Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369-387.
- (116) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D.S. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol.* **2007**, *369*, 1318-1332.
- (117) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382-W388.
- (118) Van Durme, J.; Delgado, J.; Stricher, F.; Serrano, L.; Schymkowitz, J.; Rousseau, F. A Graphical Interface for the FoldX Force Field. *Bioinformatics* **2011**, *27*, 1711-1712.
- (119) Christensen, N.J.; Kepp, K.P. Accurate Stabilities of Laccase Mutants Predicted With a Modified FoldX Protocol. *J. Chem. Inf. Model.* **2012**, *52*, 3028-3042.
- (120) Christensen, N.J.; Kepp, K.P. Stability Mechanisms of Laccase Isoforms Using a Modified FoldX Protocol Applicable to Widely Different Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3210-3223.
- (121) Dehouck, Y.; Kwasigroch, J. M.; Rooman, M.; Gilis, D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* **2013**, *41*, W333-W339.
- (122) Dehouck, Y.; Gilis, D.; Rooman, M. A new generation of statistical potentials for proteins. *Biophys. J.* **2006**, *90*, 4010-4017.
- (123) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25*, 2537-2543.
- (124) Brinda, K.V., Vishveshwara, S. A network representation of protein structures: Implications for protein stability. *Biophys. J.* **2005**, *89*, 4159-4170.

- (125) Vijayabaskar, M.S., and Vishveshwara, S. Interaction energy based protein structure networks. *Biophys. J.* **2010**, *99*, 3704-3715.
- (126) Sethi, A., Eargle, J., Black, A.A., Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6620-6625.
- (127) Stetz, G., and Verkhivker, G. M. Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: A community-hopping model of allosteric regulation and communication. *Plos Comput. Biol.* **2017**, *13*, e1005299.
- (128) Stetz, G., and Verkhivker, G. M. Dancing through life: Molecular dynamics simulations and network-centric modeling of allosteric mechanisms in Hsp70 and Hsp110 chaperone proteins. *PLoS One* **2015**, *10*, e0143752.61.
- (129) Astl, L.; Tse, A.; Verkhivker, G. M., Interrogating Regulatory Mechanisms in Signaling Proteins by Allosteric Inhibitors and Activators: A Dynamic View Through the Lens of Residue Interaction Networks. *Adv Exp Med Biol* **2019**, *1163*, 187-223.
- (130) Lange, O.F., Grubmuller, H. Generalized correlation for biomolecular dynamics. *Proteins*, **2006**, *62*, 1053–1061.
- (131) Lindahl, E., Hess, B., and Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306-317.
- (132) Palermo, G., Chen, J. S., Ricci, C. G., Rivalta, I., Jinek, M., Batista, V. S., Doudna, J. A., McCammon, J. A. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* **2018**, *51*.
- (133) Palermo, G.; Casalino, L.; Magistrato, A.; Andrew McCammon, J. Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *J. Struct. Biol.* **2019**, *206*, 267-279.

- (134) East, K. W.; Newton, J. C.; Morzan, U. N.; Narkhede, Y. B.; Acharya, A.; Skeens, E.; Jogl, G.; Batista, V. S.; Palermo, G.; Lisi, G. P., Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 1348-1358.
- (135) Nierzwicki, Ł., Arantes, P.R., Saha, A., Palermo, G. Establishing the allosteric mechanism in CRISPR-Cas9. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1503.
- (136) Martin, A. J.; Vidotto, M.; Boscariol, F.; Di Domenico, T.; Walsh, I.; Tosatto, S. C. RING: Networking Interacting Residues, Evolutionary Information and Energetics in Protein Structures. *Bioinformatics* **2011**, *27*, 2003-2005.
- (137) Piovesan, D.; Minervini, G.; Tosatto, S. C. The RING 2.0 Web Server for High Quality Residue Interaction Networks. *Nucleic Acids Res.* **2016**, *44*, W367-W374.
- (138) Floyd, R.W. Algorithm 97: Shortest path. *Commun. A.C.M.* **1962**, *5*, 345.
- (139) Hagberg, A.A.; Schult, D.A.; Swart, P.J. Exploring network structure, dynamics, and function using NetworkX, in : G. Varoquaux, T. Vaught, J. Millman (Eds.), Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, **2008**, pp. 11-15.
- (140) Girvan, M.; Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7821-7826.
- (141) Newman, M. E. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2006**, *74*, 036104.
- (142) Newman, M. E. Modularity and Community Structure in Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8577-8582.
- (143) Astl, L.; Verkhivker, G. M. Atomistic modeling of the ABL kinase regulation by allosteric modulators using structural perturbation analysis and community-based network reconstruction of allosteric communications. *J. Chem. Theory Comput.* **2019**, *15*, 3362-3380.

- (144) Astl, L.; Verkhivker, G.M. Dynamic view of allosteric regulation in the Hsp70 chaperones by J-Domain cochaperone and post-translational modifications: Computational analysis of Hsp70 mechanisms by exploring conformational landscapes and residue interaction networks. *J. Chem. Inf. Model.* **2020**, *60*, 1614-1631.
- (145) Verkhivker, G.M.; Agajanian, S.; Oztas, D.Y.; Gupta, G. Allosteric Control of Structural Mimicry and Mutational Escape in the SARS-CoV-2 Spike Protein Complexes with the ACE2 Decoys and Miniprotein Inhibitors: A Network-Based Approach for Mutational Profiling of Binding and Signaling. *J. Chem. Inf. Model.* **2021**, *61*, 5172-5191.
- (146) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **2021**, *30* (1), 70-82.
- (147) Jenik, M.; Parra, R. G.; Radusky, L. G.; Turjanski, A.; Wolynes, P. G.; Ferreira, D. U. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res* **2012**, *40* (Web Server issue), W348-351.
- (148) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M. Y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreira, D. U. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res* **2016**, *44* (W1), W356-360.
- (149) Freiburger, M. I.; Wolynes, P. G.; Ferreira, D. U.; Fuxreiter, M. Frustration in Fuzzy Protein Complexes Leads to Interaction Versatility. *J Phys Chem B* **2021**, *125* (10), 2513-2520.
- (150) Gianni, S.; Freiburger, M. I.; Jemth, P.; Ferreira, D. U.; Wolynes, P. G.; Fuxreiter, M. Fuzziness and Frustration in the Energy Landscape of Protein Folding, Function, and Assembly. *Acc Chem Res* **2021**, *54* (5), 1251-1259.

- (151) Guzovsky, A. B.; Schafer, N. P.; Wolynes, P. G.; Ferreira, D. U. Localization of Energetic Frustration in Proteins. *Methods Mol Biol* **2022**, 2376, 387-398.
- (152) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N., Structural and energetic basis of allostery. *Annu Rev Biophys* **2012**, 41, 585-609.
- (153) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014**, 508, 331-339.
- (154) White, J.T.; Li, J.; Grasso, E.; Wrabl, J.O.; Hilser VJ. Ensemble allosteric model: energetic frustration within the intrinsically disordered glucocorticoid receptor. *Philos Trans R Soc Lond B Biol Sci.* **2018**, 373, 20170175.
- (155) Li, J.; White, J.T.; Saavedra, H.; Wrabl, J.O.; Motlagh, H.N.; Liu, K.; Sowers, J.; Schroer, T.A.; Thompson, E.B.; Hilser, V.J. Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *Elife* **2017**, 6, e30688.
- (156) Leander, M.; Yuan, Y.; Meger, A.; Cui, Q.; Raman, S. Functional plasticity and evolutionary adaptation of allosteric regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 117, 25445-25454.
- (157) Wodak, S. J., Paci, E., Dokholyan, N. V., Berezovsky, I. N., Horovitz, A., Li, J., Hilser, V. J., Bahar, I., Karanicolas, J., Stock, G., et al. Allostery in its many disguises: From theory to applications. *Structure* **2019**, 27, 566-578.
- (158) Liang, Z.; Verkhivker, G. M.; Hu, G. Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Brief. Bioinform.* **2020**, 21, 815-835.

For Table of Contents Use Only