Supporting Information

Network inference from the inverse covariance matrix of dihedral angles

Using dihedral angles enables us to easily account for sidechain dynamics avoid the artifacts introduced by structural alignment\(^ {37}\). While protein dihedral angles have steep energy wells, we make the simplifying assumption that each dihedral angle has a harmonic potential energy function centered around one preferred orientation. This is only reasonable for short, picosecond-timescales, which is usually faster than the timescale for backbone dihedral flipping\(^ {63}\). Moreover, let’s assume that dihedral angle pairs only have linear coupling, which neglects nonlinear interactions and many-body interactions\(^ {28,39,40}\).

We can describe these linearly coupled dihedral angles as the Hessian matrix, \( H \), where \( H_{ij} \) is the coupling constant between a pair of dihedrals, \( i, j \). A positive coupling indicates rotation in the same direction. A negative coupling indicates rotation in opposite directions.

The harmonic coupling assumption means that \( H_{ij} \) goes as \( \text{energy radians}^2 \). Radians are a unitless measure, so \( H_{ij} \) has units \( J = \text{Nm} \), but we find it easier to use \( J/\text{rad}^2 \). Some combination of angular displacements from equilibrium, \( \theta \), then produces the torque vector \( f = -H \theta \). The potential energy difference from that at equilibrium is \( \Delta U = f \cdot d \theta = \frac{1}{2} \theta^T H \theta \).

We can also use the Boltzmann relationship to describe the probability of a particular configuration in terms of angular displacements, \( \theta \), as \( p(\theta) \sim \exp[-\Delta U/k_b T] \). The exponent is \( -\theta^T H \theta / (2k_b T) \) and can be rearranged to \( -\frac{1}{2} \theta^T (\frac{1}{k_b T} H) \theta \). This yields the probability distribution function in terms of \( H \) (equation 1).

\[
p[\theta] = \exp[-\frac{1}{2} \theta^T (\frac{1}{k_b T} H) \theta] \tag{1}
\]

By recognizing the form of a multivariate Gaussian distribution with covariance matrix \( C \) (equation 2), we can see the inverse relationship between the Hessian and covariance matrices (3). This derivation mirrors the one for anisotropic elastic network models\(^ {26,27}\), which use displacements in Cartesian space, instead of an internal coordinate system of dihedral angles.

\[
p[\theta] = \exp[-\frac{1}{2} \theta^T C^{-1} \theta] \tag{2}
\]

\[
\frac{1}{k_b T} H = C^{-1} \tag{3}
\]

Selected examples of the covariance matrix and its inverse

To demonstrate that the different patterns found between the two matrices apply to multiple MD trajectories, we show the covariance matrix diagonally across from its inverse for FimH (Supplementary Figure 1), Siglec-8 (Supplementary Figure 2), and the SARS-CoV-2 spike protein RBD-SD1 domains (Supplementary Figure 3). For each, we set the color scale maximum to the 97\(^{th}\) percentile. We show these in the same format as in Figure 1.

We highlight the backbone-backbone \( \psi-\psi \) interaction in blue, and the sidechain-sidechain \( \chi_1-\chi_1 \) interaction in red. The covariance and mutual information (Supplementary Figure 4) matrices have stronger \( \chi_1-\chi_1 \) interactions than \( \psi-\psi \). This is reversed for inverse covariance matrices.

Thresholding for visualization

While we do not use thresholds for comparing inferred interactions, we do use thresholds to visualize networks as the adjacency matrix and on the protein. In the matrix visualizations, we set the color scale maximum to the 97\(^{th}\) percentile. However, to better illustrate the pattern at lower values in the correlation matrix in Figure 2A, we also show a lower color scale maximum set to the 95\(^{th}\) in Figure S5. In contrast,
Supplementary Figure 1. The covariance matrix and its inverse show different patterns for FimH. We show that the inverse covariance matrix resembles the contact map for four representative examples. a, Wild-type FimH\textsubscript{L} (PDBID 4AUU) in 20ns and b, 200ns simulations for a different replicate than the one in Figure 1B. c, Arg60Pro mutant FimH\textsubscript{L} (PDBID 5MCA) in a 20ns simulation. d, FimH\textsubscript{2} (PDBID 4XOD), which has both the lectin and pilin domains in a 200ns simulation. The two-domain structure is visually apparent in the inverse covariance matrix. As in Figure 1B, we show the covariance matrix in the top left triangle and the inverse of the covariance matrix in the bottom right triangle. For each dataset, we set the colorscale maximum to the 97\textsuperscript{th} percentile.
Supplementary Figure 2. The covariance matrix and its inverse show different patterns for Siglec-8. We show that the inverse covariance matrix resembles the contact map for representative examples in the apo and holo states. Two replicates of Siglec-8 a, apo (without ligand, PDBID 2N7A) and b, holo (bound to 6’S-sLe³, PDBID 2N7B) See Supplementary Figure 1 for description.
Supplementary Figure 3. The covariance matrix and its inverse show different patterns for SARS-CoV-2. We show that the inverse covariance matrix resembles the contact map for representative examples of RBD-SD1 domains in the a, up, b, down, c, and off states. RBD-SD1 domains in the up and down states are from chains A and B of PDBID 6VSB. The off state is chain A of PDBID 6VXX. See Supplementary Figure 1 for description.
for the inverse covariance matrix, showing lower values does not have a profound effect, since the contact map pattern is formed by inferred interactions with high values.

We also show how thresholds affect the Jaccard similarity. In Figure 2B, we used a threshold of 97th percentile to create a mask where every edge above the threshold is set to 1, and edges below the threshold are set to 0. Here we show masks for thresholds at the 50, 95, 97, and 99th percentiles, for the covariance, mutual information, and inverse covariance matrices (Figure S8). In Figure S7, we show that for strong interactions, the inverse covariance matrix has high similarity across replicates, while the other matrices show decreasing similarities. While other matrices have higher similarity around the 50th percentile, the masks in Figure S8 show the extremely large number of edges included at these low thresholds.

The banded pattern indicates edges connecting one dihedral angle to many others, and an excessive number of these edges produces hairball networks. Thus, for the covariance, correlation, and mutual information matrices, there is a tradeoff where low thresholds with higher similarity produce hairball networks. In contrast, for the inverse covariance network, the high thresholds that produce higher similarity select for stronger interactions, producing physically interpretable networks that resemble the contact map.
Supplementary Figure 4. Mutual information matrix shows sidechain-sidechain interactions are stronger than backbone-backbone interactions. We show data for wild-type FimH. Starting with three replicate simulations of 200ns each, we show the mutual information calculated from a, transition state analysis with a core of 90 degrees and b, from a histogram based approach. For each pair of angles, we define two sets of unequally spaced bins at the deciles for each individual angle. These two sets of bins are used to construct a 2 dimensional histogram. c, We then show six replicate simulations of 20ns. The first three are truncated versions of the longer simulations.
Supplementary Figure 5. Weaker patterns in the backbone correlation and sidechain inverse covariance matrices. a, We show the covariance, correlation, and inverse covariance data from Figure 2 with the colormap set to the 95th percentile to show weaker interactions. Beneath the banded pattern in the correlation matrix, there is a weaker contact map pattern. b, Corresponding data for $\chi_1 - \chi_1$ interactions shows an extremely faint pattern that may resemble a contact in the inverse covariance matrix.
Supplementary Figure 6. The inverse covariance matrix has a “contact map”-like network at multiple thresholds. We show the covariance (top left) and inverse covariance (bottom right) at the a, 90th, b, 95th, and c, 98th percentiles to complement the 97th percentile threshold in Figure 3. We use the same representation scheme as in Figure 3.
Supplementary Figure 7. Similarity for covariance, correlation, inverse covariance, and mutual information. We calculate the Jaccard similarity for the networks defined by these four methods at different thresholds. We do this for the backbone-backbone and sidechain-sidechain interactions, and also for the entire protein. The vertical line indicates the 97th percentile used in Figure 2B. See Supplementary Figure 8 for adjacency matrix representations at representative thresholds.
Supplementary Figure 8. Representative thresholds for defining unweighted networks used to calculate the Jaccard Index (JI) for the inverse covariance, covariance, and mutual information matrices. We directly compare two replicates in the top left and bottom right triangles. Edges above the threshold in both replicates are shown in grey, while those only present in one replicate are shown in red. For each replicate pair, we list JI for $\psi - \psi$ (blue box) and $\chi_1 - \chi_1$ (yellow box) submatrices.
Supplementary Figure 9. Hierarchy of interaction strengths suggest a multilayer network. **a**, For qualitatively different interaction types, we show the distribution of interaction strengths. We compare interactions within the same residue for different dihedral angles [i, i], as well as interactions between different residues for the same dihedral angles [i, j]. The backbone label indicates the collapse of backbone layers, as calculated from the mean interaction strength for \( \phi - \psi \), \( \phi - \psi \), \( \phi - \psi \), and \( \psi - \phi \) interactions. The box-and-whisker plots mark the 5, 25, 50, 75, and 95th percentiles, with the median in blue, and outliers in grey. The Cys3-Cys-44 disulfide bond is shown with a star in dark blue. **b**, For the collapsed backbone interactions, we show the relationship between edge strength and \( C\alpha-C\alpha \) distance. Stronger interactions are associated with smaller distance. **c**, We show distributions of interaction strengths for residue pairs with \( C\alpha-C\alpha \) distances that are far apart (\( \geq 20\AA \)) and close together (\( \leq 8\AA \)), residue pairs with backbone hydrogen bonds, and neighbors on the primary sequence (i, i+1).
Supplementary Figure 10. Sidechain interactions become weaker further away from the backbone. Grey scale indicates absolute value of the interaction strength for the inverse covariance for one replicate of wild type FimH. **a,** Backbone dihedral ($\psi$) interactions with a distal sidechain dihedral ($\chi_2$) is weaker than **b,** interactions with a more proximal one ($\chi_1$). $\psi - \chi_1$ interactions still retain the contact map pattern. **c,** While interactions between distal sidechains $\chi_2 - \chi_2$ do not have a clear pattern, **d,** yet, there is still a slight pattern between proximal and distal sidechains ($\chi_1 - \chi_2$).
Supplementary Figure 11. Comparison of inferred networks for wild type and mutant FimH\_L using all inferred edges. For wild type - mutant, we show edges stronger in the wild type in blue, and those for the mutant in red (colorbar). We show two versions: a, without thresholding and b, requiring differences to be larger than twice the standard deviation of technical replicates within each group. We do not apply any filters based on distance, secondary structure, or other structural information.
Supplementary Figure 12. Steric hindrance as a potential mechanism for CC’ loop stability in apo Siglec-8. 

a. Illustration of apo Siglec-8 highlighting the CC’ loop (orange) and the carbonyl O atoms of Pro57 and Asp60 (black), which are hypothesized to be hydrogen bond acceptors for Arg70 (blue). We show residues Tyr58 and Gln59 (red) at the tip of the loop, between the hydrogen bond acceptors. 
b. Inferred networks for apo and holo Siglec-8 showed interactions within the CC’ loop, between the edges of the CC’ loop and other residues, but not interactions with Arg70. We show the average network for all 20 structures from the ensemble. 
c. Average variance of dihedral angles for the backbone and sidechains of the CC’ loop for apo Siglec-8. In 50ns MD simulations of structures 3 and 5, Arg70 formed hydrogen bonds with Pro57 and Asp60, while it did not in structure 1. Tyr58 and Gln59 in the CC’ loop have larger dihedral fluctuations for structures 3 and 5 (red and pink) than for 1 (blue). 
d. Visualization of the dynamics as overlaid 1ns snapshots show larger fluctuations for Arg70 in structure 1, while structures 3 and 5 have larger fluctuations for Tyr58 and Gln59.
Supplementary Figure 13. Impact of ligand-binding and reducing the Cys31-Cys91 disulfide bond on Siglec-8. Difference in inferred network interactions for a, apo vs holo states, and b, the holo state with and without the disulfide bond. Due to the difference in inferred interaction strength at the disulfide bond between apo and holo Siglec-8, we also compared c, the apo state with the disulfide bond intact vs the holo state without the disulfide bond. We use the same representation scheme as in Figure 4. Differences in the backbone-backbone interactions are shown in the top left in dots; sidechain-sidechain interactions in the bottom right as crosses. On the adjacency matrix, we show all differences between groups larger than 2σ, but we only draw differences larger than the 97th threshold for interaction strength on the protein structure. We show the secondary structure on the top and right borders. We show landmarks on the bottom and left borders: CC’ loop (orange), GG’ loop (purple), and the evolutionarily conserved Asp90-Cys91-Ser92 motif (green).
Supplementary Figure 14. Network inference for the trimeric spike protein. a, Backbone-backbone interactions for the trimer, with $\psi-\psi$ interactions highlighted by the blue box. The first protomer is shown with a purple box, to highlight the pattern that repeats three times within the blue box. The other backbone-backbone interactions also show a pattern that repeats three times. b, Zoomed-in view of the first protomer’s $\psi-\psi$ interactions shows a contact-map pattern.
**Supplementary Figure 15. Impact of up, down, or off state for the SARS-CoV2-spike protein RBD-SD1 domains.** In the up state (PDBID 6VSB, chain A), the RBD is accessible to bind human ACE2 for viral attachment. The neighboring protomers (chains B and C) have hidden RBDs. The off state (PDBID 6VXX, chain A) comes from a homotrimer with 3-fold rotational symmetry, where all RBDs are hidden. Difference in inferred network interactions a, among the down and off states, b, and with the up state. We use the same representation scheme as in Figure 4 and Figure S13.