

RESEARCH ARTICLE

A new approach for extracting information from protein dynamics

Jenny Liu^{1,2}  | Luís A. N. Amaral^{3,4,5}  | Sinan Keten^{1,2} 

¹Department of Mechanical Engineering, Northwestern University, Evanston, Illinois, USA

²Department of Civil and Environmental Engineering, Northwestern University, Evanston, USA

³Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, USA

⁴Department of Physics and Astronomy, Northwestern University, Evanston, Illinois, USA

⁵Department of Medicine, Northwestern University Medical School, Chicago, Illinois, USA

Correspondence

Sinan Keten, Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA.
Email: s-keten@northwestern.edu

Funding information

National Institutes of Health, Grant/Award Number: T32GM008152; National Science Foundation, Grant/Award Number: 2034584 1764421-01; Northwestern University; Office of Naval Research, Grant/Award Number: N00014163175 N000141512701; Paul and Daisy Soros Fellowships for New Americans; Simons Foundation, Grant/Award Number: 597491-01

Abstract

Increased ability to predict protein structures is moving research focus towards understanding protein dynamics. A promising approach is to represent protein dynamics through networks and take advantage of well-developed methods from network science. Most studies build protein dynamics networks from correlation measures, an approach that only works under very specific conditions, instead of the more robust inverse approach. Thus, we apply the inverse approach to the dynamics of protein dihedral angles, a system of internal coordinates, to avoid structural alignment. Using the well-characterized adhesion protein, FimH, we show that our method identifies networks that are physically interpretable, robust, and relevant to the allosteric pathway sites. We further use our approach to detect dynamical differences, despite structural similarity, for Siglec-8 in the immune system, and the SARS-CoV-2 spike protein. Our study demonstrates that using the inverse approach to extract a network from protein dynamics yields important biophysical insights.

KEYWORDS

fimbrial adhesins, molecular dynamics simulation, protein, SARS-CoV-2, sialic acid binding immunoglobulin-like lectins

1 | INTRODUCTION

Advances in experimental structure determination,^{1,2} computational structure prediction,³ and molecular dynamics (MDs) simulations⁴ have set the stage for high-throughput characterization of protein dynamics with atomic resolution. Combined with increased computational power, these advances have led to rapidly increasing numbers

of longer MD simulations for larger macromolecular systems.⁵ As a result, large datasets of MD trajectories are available from individual research labs⁵ and repositories such as MoDEL,⁶ Dynameomics,⁷ Dryad, NoMaD, and MolSSI.⁸

This wealth of MD trajectory data creates opportunities for expanding our understanding of protein dynamics and function. While snapshots from MD trajectories contain information about low energy states and can be used to identify conformational changes, some physical phenomena—such as dynamic allostery in proteins—may be

Luís A. N. Amaral and Sinan Keten contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.

better characterized by dynamics over the course of the trajectory.⁹ MD simulations can capture differences among protein variants, the impact of mutations, and modulation by small molecule binding at spatiotemporal resolutions that are difficult, or even impossible, to obtain experimentally.⁴

Taking advantage of this growing wealth of MD trajectory data will require the development of robust methods for automated analysis. A common strategy for analyzing dynamics data involves creating a network by directly calculating contact times and interaction energies.^{10–13} An alternative strategy used in network science aims to identify the underlying interactions of multicomponent systems by inferring a network structure from dynamics.¹⁴ When compared with interaction energy networks, building networks from dynamics is typically less computationally expensive and avoids less rigorous modeling of water or entropic contributions to the free energy.¹⁵ Identifying a network structure makes it possible to apply network analysis tools in order to uncover emergent properties such as densely connected communities,¹⁶ hotspots with many edges,¹⁷ and paths connecting active sites and allosteric regulatory sites in distant protein regions.^{11,18}

In the study of proteins, the typical approach for constructing networks from protein MD simulations has made use of correlation measures that quantify how different protein regions “move together.” These include a variety of methods that use linear¹⁹ and non-linear^{17,20–24} correlation measures. Yet, a rigorous mathematical analysis demonstrates that inferring the network structure by solving the inverse problem for a system that could produce the observed correlations is a more accurate approach than using the correlations directly, as described in Nguyen et al.,¹⁴ a review paper that describes the statistical mechanics underpinnings and applications in various fields. In brief, solving the inverse problem means working backwards from the observed data to solve for the parameters of a model that could have generated the data. The most straightforward form of solving the inverse problem is simply calculating the inverse of the covariance matrix.¹⁴ The inverse approach has been successfully applied to study protein coevolution,²⁵ and the inverse covariance approach is used in protein elastic network models.^{26,27} Here, we apply the inverse covariance approach to MD trajectory data, as an alternative to elastic network models.

A remaining challenge is how to define the nodes in such a network representation. An approach that has been used in elastic network models is to assign a node to each C α atom.^{26,27} This choice has some appeal because the model describes “beads,” located in Cartesian coordinates, connected by linear springs.^{26,27} However, using Cartesian coordinates requires a structural alignment step that can introduce “artifacts” during hinged motion for multidomain proteins, and even for small, single-domain proteins.²⁸ Previously, we have demonstrated that an internal coordinate system using dihedral angles makes it possible to accurately localize motion that affects elements distal to the hinge in fimH.²⁹

Here, we show that network inference using inverse covariance analysis is robust across replicates and that it uncovers strong

interactions among backbone dihedrals that form a contact-map pattern. While the contact-map pattern is also seen in elastic network models for single domains with high conformational stability, we continue to see this pattern even for multidomain proteins with hinged motion when using inverse covariance analysis.

We demonstrate the value of the proposed approach by studying three physiologically significant proteins: the bacterial adhesion protein, FimH_L, the human immune adhesion protein Siglec-8,³⁰ and two domains of the SARS-CoV-2 spike protein involved in adhesion to the human ACE2 receptor.³¹ In addition to comparing the different structures of wild-type and mutant FimH_L, we are also able to detect localized structural changes due to breaking a disulfide bond *in silico*. For Siglec-8, we are able to detect differences between “apo” and “holo” states, despite their structural similarity.³⁰ For the SARS-CoV-2 spike protein, we examined the receptor binding domain (RBD) and its connecting Subdomain 1 (SD1). While the hinge region connecting RBD-SD1 is open in the “up” state and closed in the “down” state, the individual domains remain structurally similar in the “up” and “down” states.³¹ For Siglec-8 and spike RBD-SD1, which do not have large structural changes within protein domains, comparing inferred networks allowed us to identify dynamical changes and contributions to stability.

2 | MATERIALS AND METHODS

2.1 | Protein structures

We retrieved crystal structure for FimH_L wild type and mutant, Siglec-8 apo and holo, and SARS-CoV-2 spike protein from the Protein Data Bank, as detailed in Table 1. For FimH_L, we used crystal structures of the lectin domain without ligand for both the wild type and the mutant. To compare dynamics with and without the disulfide bond as a local perturbation, we used visual molecular dynamics (VMD) to define the bond or two cysteines for FimH_L. For Siglec-8, we simulated the ligand 6'S sLe^x without the 3-amino-propyl linker, which is not thought to interact with the binding pocket.³⁰

For the SARS-CoV-2 spike protein, we started from the refined structures on the CHARMM-GUI archive.^{31,32} To focus on one hinge system that is thought to be different between the down and up states, we isolated the RBD and SD1 protein subunits without the glycans. For the “up” state, we used chain A where the RBD is accessible for binding the ACE2 receptor on human cells,³¹ and for the two “down” states, we used chains B and C. For the “off” state, we used the trimeric structure with all trimers down, checked that the backbones were very close to rotationally symmetric, and selected chain A.

We prepared all systems using VMD version 1.9.3.³³ We solvated each protein with at least 16 Å of TIP3P water molecules on each side to prevent interactions with itself through the periodic boundary conditions. We added sodium and chloride ions to neutralize the system and achieve the desired salt concentration in Table 1. The natural environment for FimH is urine, so we selected 50 mM NaCl.³⁴

TABLE 1 PDB structures studied (RRID:SCR_012820)

Protein	FimH			Siglec-8		Spike RBD-SD1		
Type	Crystal			NMR		Cryo-EM		
[NaCl] (mM)	50			150		150		
State	Wild type	Mutant	Full-length	Apo	Holo	Up	Down	Off
PDBID	4AUU	5MCA	4XOD	2N7A	2N7B	6VSB	6VSB	6VXX
Resolution (Å)	1.6	1.604	1.15	—	—	3.46	3.46	2.8
Year	2012	2017	2016	2016	2016	2020	2020	2020
Protein residues	158	158	279	145		276		
Protein atoms	2360	2350	4270	2290		4286		
System atoms	32 292	31 917	60 376	42 684	50 879	80 383	89 236	78 740
System size (Angstrom)	94	87	123	97	111	125	116	115
	61	67	72	70	72	94	95	94
	59	59	72	67	67	81	81	77

Note: We studied FimH_L in the active (wild type) and inactive (Arg60Pro mutant) states; the human immune-inhibitory protein Siglec-8 in the apo and holo (6'S sLe^x-bound) states; and the SARS-CoV-2 spike protein RBD-SD1 domains in the up and down states.

TABLE 2 Details of molecular dynamics simulations

Parameter	Value
Setup	VMD 1.9.3 (RRID:SCR_004905)
Simulation engine	NAMD 2.13 (RRID:SCR_014894)
Ensemble	NPT
Temperature	300 K
Pressure	1 atm
Nonbonded interactions	Lennard-Jones potential (cutoff)
Electrostatic interactions	Particle-Mesh Ewald sum method
Forcefield	CHARMM c36 July 2018 update
Timestep	1 fs
Coordinate saved every	1 ps
	Conjugate gradient algorithm in NAMD
Energy minimization	≥10 000 steps with protein fixed ≥10 000 steps with protein free

Abbreviation: VMD, visual molecular dynamics.

2.2 | MD simulations

We performed all-atomistic MD simulations using the Nanoscale Molecular Dynamics (NAMD) software,³⁵ with the CHARMM force field.³⁶ Our NAMD simulation parameters and system details are listed in Table 2.

After observing differences in correlated protein motions between replicates, we performed three replicates of over 200 ns each for wild-type FimH_L. Due to the tradeoff between the number of replicates and simulation length, we also performed six replicates of 20 ns of FimH_L to make comparisons of wild type and mutant FimH, as well as wild-type FimH with and without the Cys3–Cys44 disulfide bond. Qualitatively, the inferred networks are similar despite a 10-fold timescale difference.

Since 20 lowest-energy structures are reported for Siglec-8, we performed a single replicate of 50 ns for each structure to compare apo and holo Siglec-8. For the spike RBD-SD1 domains, we performed six replicates of 60 ns. To determine the timescale, we simulated a few replicates for longer to see when the inferred networks became qualitatively similar. We did this visually and by comparing the distributions of the inferred coupling strengths for adjacent residues along the peptide bond against those of distant residues.

2.3 | Backbone and sidechain dihedral angle dynamics

We used dihedral angles to capture protein dynamics because dihedral angles identify localized regions responsible for the collective displacement of regions distal from the angular rotation, such as in hinged motion.³⁷ Dihedral angles are also an internal coordinate system that avoids the structure alignment step when using Cartesian coordinates, which can introduce "artifacts".²⁸ We use both backbone (ϕ, ψ) and sidechain ($\chi_1 - \chi_5$) dihedral angles, except for Ala and Gly, as previously described.²⁹ We extracted protein features with MDTraj 1.9.5.

2.4 | Inverse of the covariance matrix

In the literature, the covariance matrix is one approach used to identify protein regions with motions that are related to the motions of many other regions; in particular, it is used to identify correlated motions between distant regions in allostery.^{22,38} However, constructing networks from the covariance matrix, even with a threshold to remove weak correlations, is susceptible to induced correlations

when two nodes (e.g., A and C) are not directly connected but share a connection with a third node (B).¹⁴ Borrowing from the field of network reconstruction, we use the inverse of the covariance matrix to identify the connections and weights, or edges, between nodes.¹⁴ This approach is consistent with finding the inverse of a covariance matrix based on C α positions, which fits the Hessian matrix describing an elastic spring network with anisotropy.^{26,27} We have found that the anisotropic elastic network model has large errors when used to describe the motion of FimH, which is consistent with errors for hinge-motion described in literature.³⁹ As a result, we use dihedral angles. This approach is similar to the torsional network model (TNM) which uses equal spring constants to describe dihedral angles across the protein.⁴⁰ In contrast, the inverse of the covariance matrix uses the variances and the covariances of angles to calculate spring constants for a network of torsional springs. The nodes in our network are dihedral angles, the edges are like linearly coupled torsional springs, and the inverse of the covariance matrix is the Hessian matrix for a TNM.

To construct our network, we calculate the Moore–Penrose pseudo-inverse of the covariance matrix using both the backbone and sidechain dihedral angles. We use this approach to understand the relative contributions of backbone and sidechain dynamics to collective motion. Since the sign describes whether the angles turn in the same direction, we take the absolute value to get the interaction strength. We do not apply distance filters. While we use the 97th percentile as a value threshold for selecting strong interactions or visualizing the network on the protein, we do not use any thresholds for network comparisons. More generally, we recommend caution for applying thresholds to these networks for analysis.

For purposes of illustration, we focus on the χ_1 sidechain dihedral. However, we calculate the inverse covariance matrix with χ_{1-5} dihedrals.

2.5 | Comparing networks of inferred interactions

To identify interactions that are stronger in one protein state than another, we compare each edge. We select for large differences between groups (e.g., WT vs. mutant), relative to the variability within each group. To do this, we filter for differences larger than twice the standard deviation for each group. To compare an edge e between states a and b , each with an ensemble of m and n networks, this is $|\langle e_a \rangle_m - \langle e_b \rangle_n| > 2\sigma_{e_a}$ and $> 2\sigma_{e_b}$. We apply this rule without determining statistical significance with corrections for multiple comparisons, in order to see the full effects of comparing all interactions on the matrix. In Figure S11, we also show an example of comparisons without filtering for large differences, in order to illustrate the persistence of the contact-map pattern. We perform the network comparisons in two ways: (1) for every edge on the network (see Figure S11), (2) accounting for the multilayer structure of the network by collapsing the backbone–backbone interactions into residue–residue interactions (Figure 4). We analyzed data in python, using SciPy (RRID:SCR_008058) and custom packages.

3 | RESULTS

We present results below for these three proteins (Figure 1A). We first focus on the well-characterized allosteric protein FimH_L. Separation of the FimH_L domain from its connecting domain (bottom in all figures) is thought to induce an allosteric conformational change on the opposite end of the protein (top in figures).^{41,42} This changes the binding pocket from a state with low affinity for the ligand to one with high affinity (Figure 1B).⁴¹ While wild-type FimH_L is trapped in the high-affinity state, a single-amino acid mutation (Arg60Pro) stabilizes FimH_L in the low-affinity state.^{43,44} The mutant FimH_L is of interest because it undergoes an allostery-like conformational change upon binding mannoside ligands and has been proposed as a minimal model of allostery.⁴⁴

Like FimH_L, Siglec-8 binds a carbohydrate ligand and has an immunoglobulin-like fold with two β -sheets (Figure 1C). Functionally however, Siglec-8 binds to specific sugars found uniquely in human airway tissues to prevent autoimmunity.^{30,45} For specific binding, Siglec-8 has a surprisingly rigid binding pocket loop, leading to similar structures for the apo and holo states.³⁰ The SARS-CoV-2 spike protein RBD binds the human ACE2 receptor with a “hook” region.⁴⁶ The hook becomes accessible in the up state when the hinge between the RBD and its connector opens (Figure 1A). While the hook and interdomain hinge regions are flexible, the bulk of the RBD is thought to be structurally similar in the up and down states. Both Siglec-8 and the spike RBD-SD1 present challenges for detecting differences in inferred interactions because the protein state changes without major structural changes within domains.

3.1 | Define and validate network inference from inverse covariance analysis

Our approach for constructing a network representation of the dynamics of a given protein is comprised of three steps. In the first step, we obtain temporal dynamics for the nodes, which are the backbone (ϕ , Ψ) and sidechain (χ_{1-5}) dihedral angles for each residue.^{24,37} In the second step, we calculate the circular covariance for dihedral angles,³⁷ which can be thought of as the linearization of the interactions captured by mutual information. In the third step, we invert the covariance matrix using the Moore–Penrose pseudo-inverse to calculate the best fit for a linear coupling system that can give rise to the observed covariance matrix.¹⁴

Similar to mutual information calculations conducted on other proteins,^{17,24} we find that backbone–backbone interactions computed from the covariance are weak compared with sidechain–sidechain interactions (compare red and blue boxes in Figure 1B and the mutual information in Figure S4). In these networks, some dihedral angles have a banding pattern, suggesting long-range interactions with many other dihedral angles (Figures 1B and S1). In contrast, the inverse covariance matrix has localized and specific interactions. In addition, the stronger backbone–backbone interactions have a repeating pattern that resembles the contact map of the protein, and this pattern appears to repeat more weakly in backbone–sidechain interactions.

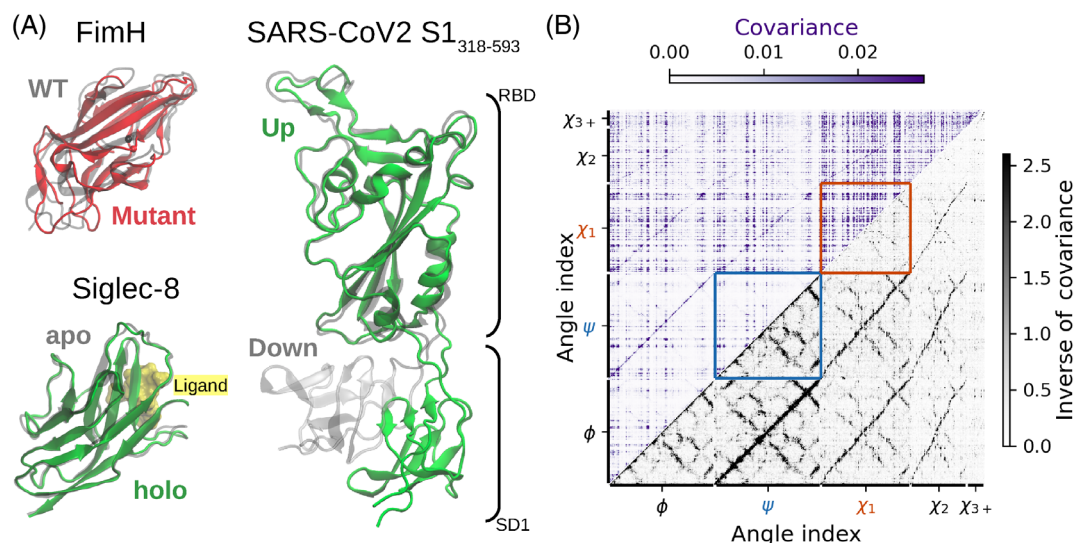


FIGURE 1 Unraveling structural properties from protein conformational dynamics. (A) Cartoon illustrating the three adhesion proteins studied here. FimH refers to the lectin domain of a bacterial adhesin found in uropathogenic *Escherichia coli* that binds mannose and undergoes a conformational change under tensile force from urine flow. Siglec-8 refers to the lectin domain of a human immune-inhibitory protein found on eosinophils and mast cells. The SARS-CoV-2 receptor binding domain (RBD) and Subdomain 1 (SD1) domains are thought to undergo a down-to-up transition that makes the RBD available to bind ACE2. For each protein, we compare pairs of states: FimH_L wild-type (PDB 4AUU) and mutant (PDB 5MCA), Siglec-8 with ligand 6'S-sLe^x (PDB 2N7B) and without lig (PDB 2N7A), and RBD-SD1 in down and up (PDB 6VSB) states. (B) Comparison of covariance analysis of the dynamics (top left) versus the inverse covariance analysis (bottom right) from the dynamics of wild-type FimH_L (see Figures S1–S3 for the other proteins). While many studies rely on the analysis of the covariance matrix, our data clearly show that the structure of the covariance matrix is dominated by artifacts (vertical and horizontal lines) which are stronger for side-chain interactions (red square for $\chi_1 - \chi_1$). In contrast, the inverse covariance matrix clearly reveals a structure reminiscent of a contact map and is dominated by backbone interactions (blue square for $\psi - \psi$)

The banding pattern in the covariance matrix is also widespread in mutual information matrices, where they have been interpreted as long-range interactions important in protein allostery.^{17,24} The large number of long-distance edges produce “hairball” networks, which led to the use of pruning algorithms^{47,48} or distance filters^{22,38,49} in prior studies, in order to make network analysis tractable. Thus, we wondered whether the long-range interactions are capturing a physical feature of the dynamics. To answer this question, we investigate the reproducibility of the covariance, correlation, and inverse covariance matrices extracted from different replicates of MD simulations.

In Figure 2A, we contrast the matrix for one replicate in the upper-diagonal with the second replicate in the lower-diagonal and quantify the similarity in Figure 2B. For replicate MD simulations, we used the same initial protein structure with randomized solvation and initial velocities. Both covariance matrices have banding patterns suggesting hotspots that interact with many residues across the protein. However, each replicate has its own banding pattern, with interaction strengths that are over 10 times greater than those found in the other replicate, indicating high variability in the networks that one would construct from replicate simulations.

Since the banding pattern is associated with dihedral angles with high variance, we also consider the correlation matrix, which normalizes the covariance matrix by the variance of each dihedral angle. It is visually apparent that the banding in the covariance matrix is not simply due to high variance because there are still bands in the

correlation matrix. While normalizing to the correlation matrix uncovers some interactions in a contact map pattern, they are weak compared with the banding pattern (see Figure S5 for a lower maximum value on the color map scale).

In contrast to the irreproducible results obtained with the covariance matrix, for the inverse of the covariance matrix, we find a pattern that is visually similar to the 12 Å contact map (Figure 2C). The diagonally symmetric contact map pattern in blue indicates similarly strong interactions for two replicates (Figure 2A). After quantifying the robustness across three replicates using the Jaccard similarity index, we find that the inverse covariance has higher similarity (59%–72% shared edges) than the covariance (8%–10%) or the correlation (13%–16%) for $\psi - \psi$ backbone interactions (Figure 2B). The $\chi_1 - \chi_1$ similarity values for inverse covariance are lower, but still higher than for the other two methods. These data clearly demonstrate that networks inferred from inverse covariance analysis are more robust than networks constructed from correlational measures.

3.2 | Inverse covariance analysis yields structural networks

Prompted by the strong visual resemblance between the inverse covariance network and the 12 Å contact map and the complete absence of this pattern in the covariance network, we wondered if

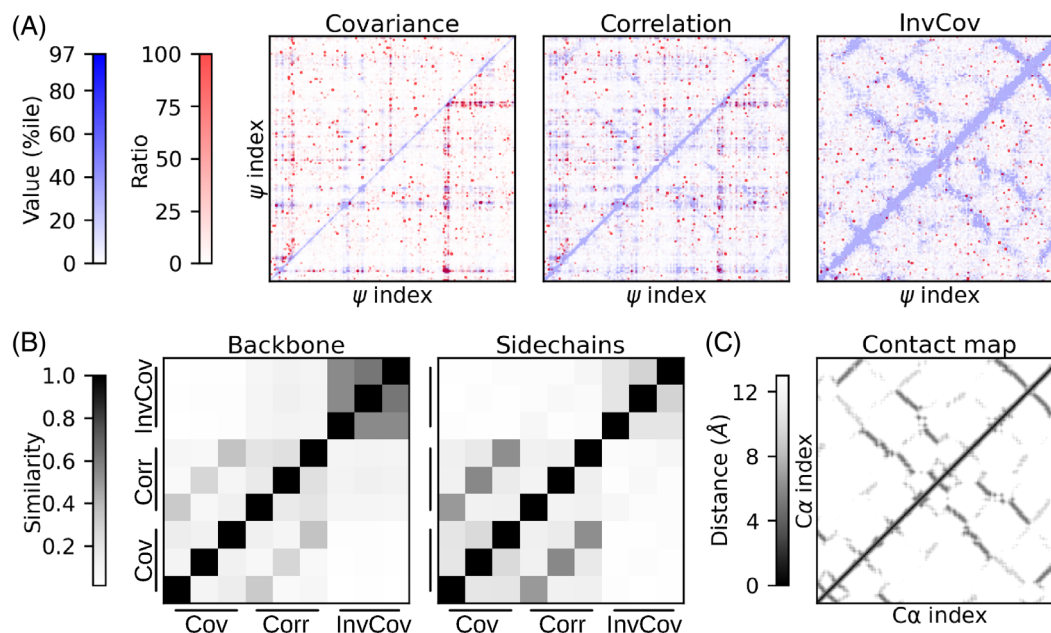


FIGURE 2 Inverse covariance matrix is robust across replicates whereas covariance and correlation matrices are not. (A) Triangular regions above and below the matrix diagonal show results from two replicates of wild-type FimH_L starting from the same protein structure. We show $\psi - \psi$ interactions. We show interaction strength in blue with a normalization for each triangular region made based on the 97th percentile of observed strengths. In red, we show the ratio for interaction strength between the two replicates. Purple indicates strong interactions that are not reproduced in the other replicates (see Figure S5 for weaker interactions visible when normalized to the 95th percentile for $\psi - \psi$ and $\chi_1 - \chi_1$). For covariance and correlation matrices, we find that backbone–backbone interactions are mostly quite weak, but the strong interactions (darker blue) vary drastically between replicates. In contrast, for the inverse covariance matrix, the strongest backbone–backbone interactions symmetric across the diagonal. (B) To evaluate the robustness of network inference, we calculate the Jaccard similarity coefficient for the covariance, correlation, and inverse covariance analyses methods across three simulation replicates. We define edges above the threshold of ≥ 97 th percentile (see Figure S7 for other thresholds). In gray scale, we show similarity separately for $\psi - \psi$ and $\chi_1 - \chi_1$ interactions. Darker gray indicates results are similar across replicates for the inverse covariance approach and much less similar for the other two methods. (C) The inverse covariance matrix resembles the contact map. We show the C α inter-residue distance from the crystal structure. Darker gray indicates shorter distance

the inverse covariance matrix could be used to identify specific physical interactions. To answer this question, we overlaid the strongest edges (≥ 97 th percentile) on the 12 Å contact map, highlighting the backbone–backbone edges as blue dots and the sidechain–sidechain edges as red crosses (Figure 3A and Figure S10).

To correct for the high variability across replicates that we previously found for the covariance networks, we averaged networks across the three replicates. Despite the averaging, the covariance network shows strong interactions across distant protein regions and is dominated by $\chi_1 - \chi_1$ interactions.

To understand how these contrasting patterns affect interpretation, we next visualize strong interactions as edges drawn on the protein structure (Figure 3B). Since drawing all edges would make the covariance network indecipherable, we only show the

$$\chi_1 - \chi_1$$

edges originating from Lys4. In contrast, the inverse covariance network uncovers edges that mostly connect physically close residues. Specifically, we find a $\chi_1 - \chi_1$ edge between Cys3 and Cys44 for the

sole disulfide bond in FimH_L, and that this edge is missing from the covariance network (red vs. green arrow in Figure 3A).

Examining the contact map pattern of the inverse covariance network in more detail, we compare edge weight with the distance between C α atoms (Figure S9). We find that for backbone–backbone interactions, the strongest interactions are between residues connected by a peptide bond, followed by hydrogen bonds within β sheets, and then nonbonding interactions.

After examining backbone–backbone interactions, we next looked at the progressively weaker interactions involving sidechains distal from the backbone (Figures 1B and S10). We find the contact map pattern is still apparent for $\phi - \chi_1$ or $\psi - \chi_1$ interactions, but becomes very weak for $\phi - \chi_2$ or $\psi - \chi_2$ interactions, and becomes indistinguishable from noise for interactions between proximal and distal sidechain dihedrals. The inverse covariance analysis thus suggests that backbone dihedral motion is most strongly coupled to nearby backbone dihedrals and has more dissipated effects on sidechain dihedrals. This relationship is consistent with how backbone motions can sterically trap or free sidechains, whereas sidechain motions are more limited in their impact on backbone motion.⁵⁰

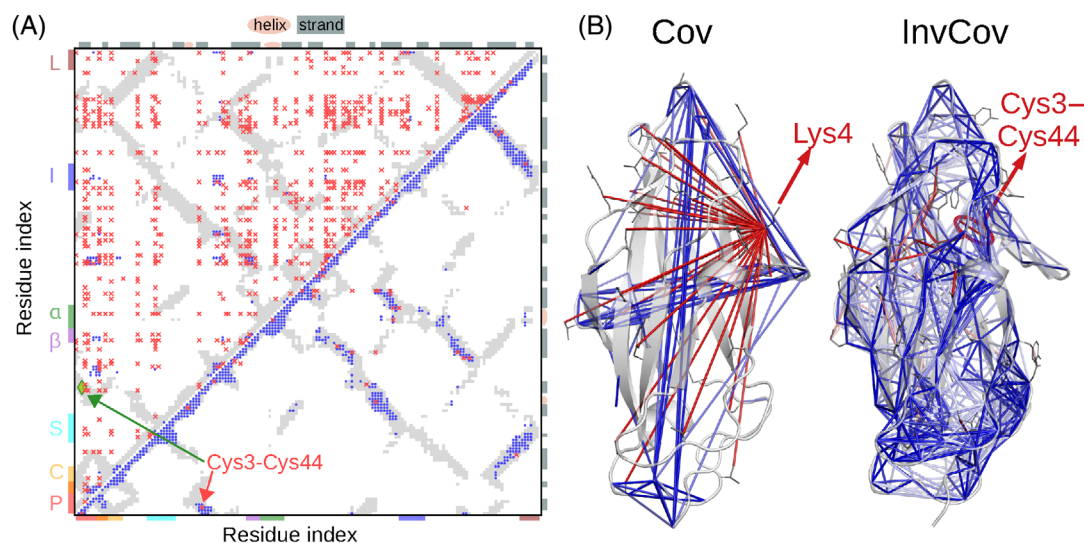


FIGURE 3 The inverse covariance matrix enables us to extract a “contact map”-like network from the protein dynamics. (A) Comparison of strong interactions identified for the covariance matrix (top left) and for the inverse covariance matrix (bottom right) of wild-type FimH. To provide context for our data, we plot the 12 Å contact map in gray within the matrix. On the top and right axes, we show helix (pink) and strand (teal) secondary structures assigned by the Dictionary of Secondary Structure of Proteins algorithm. On the left and bottom axes, we show putative allosteric pathway landmarks⁴³: pocket zipper (red), clamp segment (yellow), swing loop (cyan), β-bulge (purple), α-switch (green), insertion loop (blue), and linker loop (dark red). We only show strong edges, with the threshold set at the 97th percentile of all dihedral interactions. See Figure S6 for other cutoffs. We averaged edge weight across three replicates. The blue dots represent the average of backbone-backbone interactions by residue. The red crosses represent sidechain-sidechain interactions ($\chi_1 - \chi_1$). The inverse covariance network is predominantly backbone interactions that fall within the 12 Å contact map. There is a ≥ 99 th percentile $\chi_1 - \chi_1$ interaction: the Cys3–Cys44 disulfide bond (red arrow). However, this interaction is only 80th percentile in strength (green diamond and arrow) for the covariance matrix, which is dominated by other $\chi_1 - \chi_1$ interactions. (B) Since the covariance matrix has many long-range interactions, we only show the backbone interactions and limit the sidechain interactions to Lys4. In contrast, the inverse covariance network has mostly short-range interactions, including the disulfide bond. We show backbone interactions on the C α atoms and sidechain interactions on the fourth χ_1 atoms. We show backbone interactions in blue and sidechain interactions in red. Darker colors indicate stronger interactions

The different strengths of interactions for backbone-backbone and backbone-sidechain edges suggests that qualitatively different types of interactions have different properties. For two residues i and j , the backbone-backbone edge $\phi - \psi[i, j]$ are larger than the backbone-sidechain edge $\chi_1 - \chi_1[i, j]$, which is consistent with the physical differences between these two edge types. Moreover, most backbone-backbone edges within the same residue, $\phi - \psi[i, i]$, are stronger than backbone edges connecting to other residues, $\phi - \phi[i, j]$ and $\psi - \psi[i, j]$ (Figure S9).

3.3 | Detecting both large and small structural changes in FimH_L

3.3.1 | Conformational differences between wild type and mutant

As a way to further validate our approach, we next test if we are able to identify the well-characterized differences between wild-type FimH_L and the Arg60Pro mutant. To compare inferred networks for the wild-type and mutant proteins, we identified edges where the average difference was larger than two times the standard deviation across each group of replicates. We performed this analysis once with

the entire set of edges (Figure S11), and again with only the backbone interactions collapsed into a residue interaction network. For our comparisons and the matrix visualization of the differences, we do not apply a distance filter or a threshold for the edge-strength. However, for the visualization on the protein, we only show differences with magnitude larger than the 97th percentile.

The visualization of the differences enables us to identify several interactions stronger in either the mutant or the wild-type proteins (red or blue patches, respectively, in Figure 4A). This is consistent with the difference in initial structure (root mean squared deviation, RMSD = 3.15 Å), dihedral dynamics,²⁹ and the residues that are rearranged in the allosteric conformational change.^{43,51}

For concreteness, we focus on two regions at the edge of the protein structure that are easier to visualize: the binding pocket zipper at the top of FimH_L, and the insertion loop at the bottom. In the pocket zipper, we found much stronger interactions for the mutant protein (median: 2.8-fold, IQR: 2.0-fold to 5.2-fold), which correspond to smaller dihedral fluctuations.²⁹ On the other hand, in the insertion loop, we identified changes in interaction that were stronger in the wild type than the mutant protein (3.2-fold, 2.2-fold to 3.7-fold). Structurally, this is consistent with how the insertion loop is stabilized in the wild-type structure and exposed to solvent in the mutant protein.^{44,52} Dynamically, stronger interactions within the insertion loop

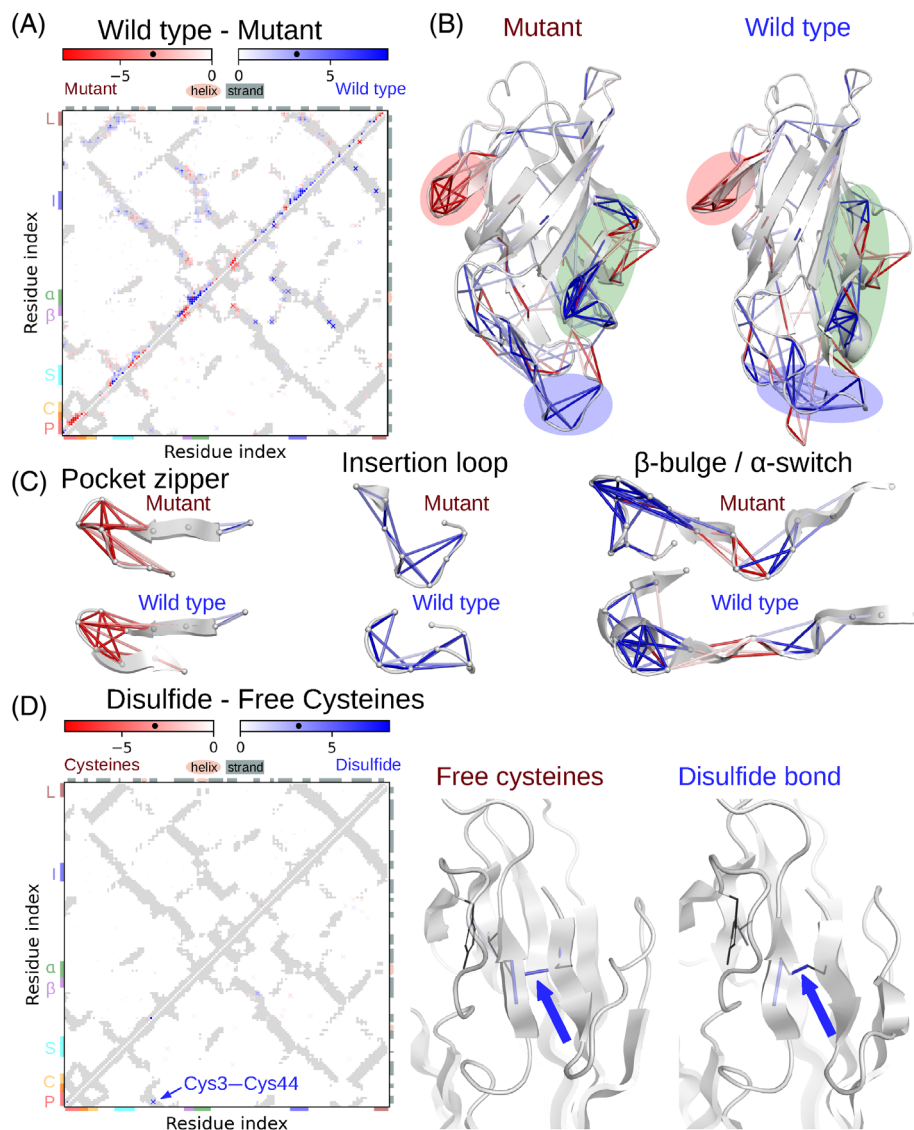


FIGURE 4 Inverse covariance analysis can detect both large and small structural changes in FimH_L. (A) Comparing inferred networks for wild-type and mutant proteins, we show differences in the backbone (top left in dots) and $\chi_1 - \chi_1$ (bottom right in crosses). As in Figure 3A, we annotate the 12 Å distance cutoff, secondary structures, and landmarks. The black dots on the color bar mark the 97th percentile in magnitude. On the adjacency matrix, we show all differences greater than 2σ . (B) On the protein, we show large differences (≥ 97 th percentile). Red shows interactions stronger for mutant FimH_L; blue for WT FimH_L. We highlight the pocket zipper (red), insertion loop (blue), and β -bulge/ α -switch (green) and (C) show isolated parts of the network. (D) Comparing wild-type FimH with the Cys3–Cys44 disulfide bond intact or reduced *in silico*. (E) The blue lines show that the Cys3–Cys44 $\chi_1 - \chi_1$ (blue arrow) and the Phe43–Cys44 backbone–backbone interactions are stronger when the disulfide bond is intact

are consistent with smaller dihedral fluctuations in the wild-type protein.²⁹

We further identify differences at the β -bulge, α -switch, and swing loop regions of the allosteric pathway, consistent with structural differences between wild-type and mutant proteins. The mutant protein has stronger interactions within the loop formed by the β -bulge (2.0-fold, 1.2-fold to 2.3-fold) and also with a nearby loop. In the wild-type protein, the loop is smoothed out into a β -strand. The wild-type protein has stronger interactions (1.8-fold, 1.5-fold to 2.6-fold) in the α -helix, compared with the 3_{10} -helix in the mutant protein, which is probably due to different hydrogen bonding patterns. In the swing loop, we again find stronger interactions in the wild-type protein (2.0-fold, 1.6-fold to 2.7-fold).

For these allosteric pathway landmarks, it is visually apparent that we detect large differences in inferred interactions when structures are closer in one state and stretched apart in the other state. Beyond these regions, there are several other regions with similarly large

changes in interaction between the wild-type and mutant proteins, shown in blue and red patches (Figure 4A).

3.3.2 | Quantifying the impact of disulfide bond reduction

We next used wild-type FimH_L to explore the impact of reducing the single disulfide bond between Cys3 and Cys44 *in silico* on fast, nanosecond-timescale dynamics. Using the inverse covariance analysis, we correctly identified the 6-fold stronger $\chi_1 - \chi_1$ interactions in the presence of the disulfide bond, which was the largest difference detected (Figure 4D). This matches our expectations because the covalent bond between the most distal atoms forming the χ_1 rotamer angle directly couples χ_1 dynamics. Together, these analyses show that the inverse covariance analysis method is sensitive to both local differences and conformational differences.

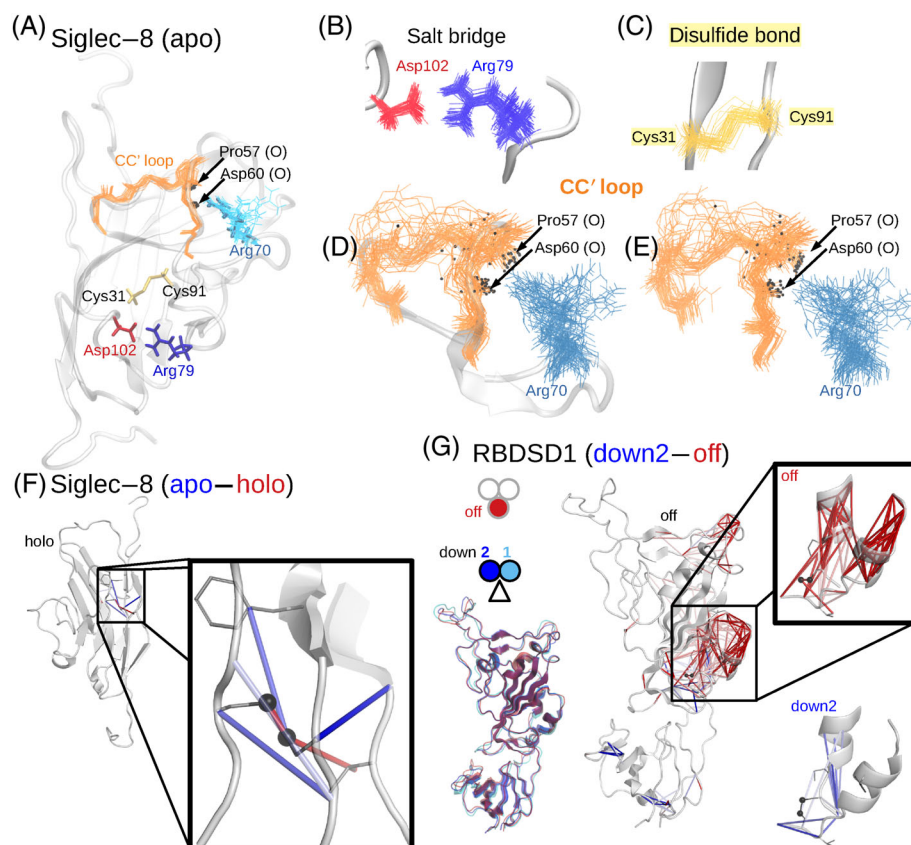


FIGURE 5 Inferred networks identify strong interactions and changes in interaction strength. (A) Illustration of Siglec-8 highlighting the CC' loop of the binding pocket in orange, Arg70 in cyan, and hypothesized hydrogen bond acceptor carbonyl oxygens (black spheres) for ensemble of 20 NMR structures. From molecular dynamic simulations of all structures, we show snapshots every 10 ns for the (B) Cys31–Cys91 disulfide bond, (C) Arg79–Asp102 salt bridge, and (D) CC' loop. Structures were aligned to the backbone atoms for residues 13–135 to exclude the N-terminus and C-terminus tails. These snapshots show the stability of the salt bridge and disulfide bond, and the surprising stability of the unstructured CC' loop. However, the hypothesized interaction between Arg70 and the CC' loop is much less stable, even with (E) structural alignment using the backbone of the CC' loop. (F) Beyond identifying strong interactions, our approach also identifies rearrangements of strong interactions without large structural changes in Siglec-8 and the SARS-CoV-2 RBD-SD1. Siglec-8 with (holo) and without (apo) ligand have similar structures. However, inverse covariance analysis reveals rearrangement of strong interactions in a region opposite the binding pocket, including the Cys31–Cys91 disulfide bond (black spheres). We show interactions stronger in holo (red) and apo (blue) on the holo structure. (G) For the SARS-CoV-2 spike protein, rotation around the RBD-SD1 hinge exposes the RBD in the up conformation (Figure 1A), while the structures with the RBD hidden are extremely similar (down1, down2, and off conformations). Comparing inferred network interactions, we detect differences near the hinge, including the Cys336–Cys361 disulfide bond (black spheres) and nearby α -helices. For these regions, we show the interactions that are stronger in the down2 (blue) and off (red) conformations on the corresponding structure. See Figure S15 for other comparisons

3.4 | Network inference detects key mechanisms involved in Siglec-8 binding

Like FimH_L, the human immune cell adhesion protein, Siglec-8, is also a lectin with an immunoglobulin-like fold with a single disulfide bond. For Siglec-8, we compare the apo (no ligand) and holo (bound to the native 6'S-sLe^x ligand) states. Due to the rigid binding pocket loop that only differs by a few sidechain rearrangements, apo and holo Siglec-8 have extremely similar structures.³⁰ Intriguingly, the rigidity of the CC' binding pocket loop in apo Siglec-8 occurs in the absence of stabilizing secondary structure motifs³⁰ and plays a major role in recognizing specific ligands in the airway to avoid autoimmunity.⁴⁵

One hypothesized mechanism for stabilizing the CC' loop in apo Siglec-8 is that the Arg70 sidechain forms hydrogen bonds with the

loop backbone at Pro57 and Asp60 (Figure 5A).³⁰ We identified Arg70–Pro57 or Arg70–Asp60 hydrogen bonds with occupancy between 10% and 33% in seven out of 20 simulations from the ensemble of NMR structures, and higher occupancy (80% and 83%) in only two simulations. This suggests the hydrogen bond is rarely present. Nonetheless, the unstructured CC' loop has surprisingly small fluctuations across the ensemble of MD simulations (Figure 5B–E).

Using our inferred interaction networks, we identified strong interactions within the CC' loop, as well as interactions from outside the loop to its hinges at Ala53 and Pro62 (Figure S12). Consistent with the hydrogen bonding analysis, we only identified strong interactions with Arg70 in a few MD simulations, which disappeared after averaging the networks across the ensemble. Counter-intuitively, in the two structures where Arg70 does form stable hydrogen bonds

with Pro57 and Asp60, the CC' loop has paradoxically larger fluctuations, especially at Tyr58 and Gln59 at the loop tip (Figure S12). Taken together, we suspect that steric hindrance plays a role in stabilizing the CC' loop internally and externally at the hinge edges.

Of note, this example demonstrates an instance where the inferred network approach provides more clarity than contact or interaction energy networks, because the inverse of the covariance matrix identifies the degrees of freedom that most strongly affect the fluctuations at the CC' loop.

After examining why the CC' loop is similar in apo and holo Siglec-8, we next compared inferred networks for the apo and holo states. The holo state has stronger $\chi_1 - \chi_1$ interactions corresponding to the Cys31–Cys91 disulfide bond (Figures 5C and S13). The disulfide bond is conserved in the siglec family and is located on the sheet of the β -sandwich opposite the binding pocket.³⁰ Nearby, we also observe other changes in interaction strength involving Asp90. Although distant from the binding site, the Asp90–Cys91–Ser92 motif in Siglec-8 is a variant of the Asn–Cys–Ser or –Thr motif that is conserved in the rest of the siglec family.⁵³ Differences in interaction strength between apo and holo states identify changes in the dynamics of this evolutionarily conserved region during ligand-binding.

In contrast, reducing the Cys31–Cys91 disulfide bond in the holo state *in silico* has a different pattern (Figure S13). Both ligand-binding and the presence of the disulfide bond correspond to stronger a $\chi_1 - \chi_1$ interaction at Cys31–Cys91. Both conditions also stabilize Cys31, indicated by decreased backbone dihedral fluctuations and increased duration within an extended secondary structure as assigned by the Dictionary of Secondary Structure of Proteins algorithm, and shorter Cys31–Cys91 C α distance. Taken together, we find that the network rearrangement that occurs with ligand-binding increases Siglec-8 conformational stability near an evolutionarily conserved disulfide bond, even though the region is not near the binding site.

3.5 | Comparison of the spike protein RBD-SD1 fragments in the up, down, and off states shows network rearrangements without large structural differences

Next, we investigated whether there are differences in the networks inferred from the “up,” “down,” and “off” states of the RBD-SD1 domains of the SARS-CoV-2 spike protein (Figure 1A). The RBD connects to SD1 (Figure 5E) via two hinge-like loops that are more flexible in the up state than in the down or off states.⁵⁴ While the up state has a different orientation around the hinge than the down and off states, they all have similar SD1 structures (C α -RMSD \leq 0.64 Å) and somewhat similar RBD structures (C α -RMSD \leq 1.55 Å).

Opening the hinge angle in the down-to-up transition is thought to make the binding site on the RBD available to attach to human ACE2.^{31,46} To focus on the RBD-SD1 hinge, we isolated these domains from the rest of the spike protein and ignored glycosylated sugars. While these simplifications limit the strength of our

conclusions into the function of the spike protein, the RBD-SD1 structures nonetheless provide a useful system for comparing dynamics in a system that initially resembles rigid-body motion around a hinge.

Since the trimeric spike protein with one exposed RBD has two hidden RBDs, we isolated one RBD-SD1 fragment in the “up” state, and two fragments in the “down” state. We obtained another RBD-SD1 domain in the “off” state from a structure with all RBDs hidden and 3-fold rotational symmetry. We first compared the RBD-SD1 fragments in the down and off conformations. Despite the structural similarity, we nevertheless detect differences in the inferred networks (Figure 5E) based on dynamics. We find the networks for the two down conformations are more similar to each other than to the networks for the off conformation (Figure S15).

We next compared RBD-SD1 fragments in the down and off conformations to the fragment in the up conformation (Figure S15). Surprisingly, we find that the differences in inferred networks among the down and up conformations are more localized, while the networks for the off conformation are distinct from the others. The localized differences are in the hook of the RBD that is exposed in the up conformation, and at an α -helix near the RBD-SD1 hinge. Intriguingly, we also identified differences in the structural orientation and the inferred network interaction at the Cys336–Cys361 disulfide bond. In this region, one down conformation is more similar to the up conformation, whereas the other down conformation resembles the off conformation (Figure 5E).

As a proof-of-concept, we investigated whether it is possible to infer an interaction network for the trimeric SARS-CoV-2 spike protein, even though it is much larger than the RBD-SD1 domains. However, the amount of data required for network inference using the inverse covariance method scales faster than the number of residues. As a result, our approach required more data than was available from the two sets of publicly accessible simulations for which the up and down states are labeled.^{55,56} Instead, we chose the longer simulations that explore the transition among the down, up, and open states published by the Bowman lab.⁵⁷ We randomly concatenated the short simulations to create progressively larger datasets for calculating the inverse covariance matrix. As the amount of data increased, the distribution of coupling strengths between neighboring residues along the peptide backbone also became distinct from those between distant residues. Using 100 000 snapshots, we were able to infer an interaction network for the 3363 residues of the spike protein (Figure S14). The number of snapshots is an order of magnitude larger than those currently available for labeled states. Thus, it may be feasible to compare states for the entire S1/S2 complex of the spike protein if there is sufficient data, or by choosing a less data-intensive method for solving the inverse problem.

4 | DISCUSSION

We identify some of the shortcomings of correlation-based approaches for network inference from protein dynamics using the

covariance, correlation, and mutual information matrices. These networks have low reproducibility among replicates and exhibit long-range connections that are difficult to tie to physical explanations. To address these shortcomings, we use the inverse of the covariance matrix. This is a well-established technique from network inference¹⁴ for solving the inverse problem for a system that can produce the observed correlated dynamics. Our approach builds networks where each node is a dihedral angle, including both the backbone (ϕ , ψ) and the sidechains (χ_{1-5}), and edges are inferred from the coupling interactions between angles. We chose the internal coordinate system of dihedral angles³⁷ to easily include sidechain dynamics, localize hinges that drive distal dynamics,²⁹ and to avoid the alignment step in Cartesian coordinates that introduces “artifacts” in hinged motion.³⁹

Using the inverse covariance approach, we detected differences in conformation, subtle differences between protein states without large conformational changes, and localized perturbations in the structure of biomedically important proteins. The inverse covariance networks capture a hierarchy of interactions that resemble the qualitatively different types of interactions, suggesting a multilayer network structure. The strongest edges connect dihedral angles with covalently bonded atoms, with weaker interactions for greater distances. Moreover, the contact map-like pattern found in backbone-backbone interactions is repeated more weakly in backbone-sidechain and sidechain-sidechain interactions. This hierarchy of inferred interactions is consistent with the view that smaller backbone rearrangements are related to larger sidechain motions.⁵⁰ Since stronger coupling leads to smaller fluctuations and vice versa, a comparison of coupling strengths may capture shifts in ensembles that have similar average structure, as shown for Siglec-8 and the hinge region of SARS-CoV-2. The inverse covariance approach captured the role that disulfide bond play on the dynamics of these adhesion proteins, in the sense that disulfide removal strongly impacts coupling strengths and major differences in networks tend to occur at disulfide bonds when comparing different states of these proteins. More generally, comparing inferred network properties may be useful in dynamic allostery^{9,58} and thermostable variants.⁵⁹ Specifically, a multilayer network approach may allow the use analytical tools to capture properties lost by flattening networks.⁶⁰

Our results suggest that solving the inverse problem uncovers the underlying interactions that ultimately drive protein dynamics, but are not well-captured by cataloging the observed correlated motions or comparing static structures. Notably, inverting the covariance matrix is the simplest of a variety of tools available for network inference from dynamics used in network science.^{14,61} While the simplicity of inverting the covariance matrix increases accessibility, there are some obvious limitations.¹⁴ We calculate the circular covariance matrix on dihedral angle distributions that are multimodal. The inverse of the covariance matrix is analogous to linearly coupled torsional springs, which do not represent the complexity of atomic interactions within a protein. Moreover, network inference by inverting the covariance matrix requires a large amount of data.¹⁴ Our work establishes a baseline approach, which can be easily built upon by incorporating more sophisticated^{14,18}—and yet more involved—approaches that better

describe dihedral distributions²⁵ or account for nonlinear interactions.⁶²

Despite these limitations, our approach yielded significant insights for three adhesion proteins. Comparing the networks inferred for two protein states at a time, we were able to tie differences in inferred network structure to structural and dynamical differences. For FimH_L, a comparison of inferred networks for the wild-type and mutant proteins identifies protein regions with conformational changes consistent with the allosteric pathway sites.⁴³ For Siglec-8 and the SARS-CoV-2 RBD-SD1 construct, we were able to detect network rearrangements despite the similar structures of Siglec-8 in the apo and holo states, and of the individual RBD and SD1 domains in the up and down states. In Siglec-8, we were also able to use strong interactions identified by the network to propose a new mechanism for stabilizing an unstructured loop in the binding pocket. This serves as an example where the network inference approach has an advantage over contact and interaction energy networks. Taken together, our results show that the network inference approach can identify protein regions of interest based on dynamical differences that are rooted in physically interpretable interactions.

ACKNOWLEDGMENTS

The authors thank Martin Gerlach and Kerim Dansuk for helpful conversations. Jenny Liu thanks the Paul and Daisy Soros Fellowship, the Northwestern Quest High Performance Computing Cluster, and the National Institute of Health T32GM008152. This project was also supported by the Office of Naval Research N00014163175 and N000141512701 (Sinan Keten), the National Science Foundation 2034584 (Sinan Keten) and 1764421-01 (Luís A. N. Amaral), and the Simons Foundation 597491-01 (Luís A. N. Amaral).

CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. An example code for the inverse covariance approach is available on github at <https://github.com/keten-group>.

ORCID

Jenny Liu  <https://orcid.org/0000-0002-2867-8687>

Luís A. N. Amaral  <https://orcid.org/0000-0002-3762-789X>

Sinan Keten  <https://orcid.org/0000-0003-2203-1425>

REFERENCES

1. Levitt M. Growth of novel protein structural data. *Proc Natl Acad Sci U S A*. 2007;104:3183-3188.
2. Terwilliger TC, Stuart D, Yokoyama S. Lessons from structural genomics. *Annu Rev Biophys*. 2009;38:371-383.
3. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589.
4. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. 2018;99(6):1129-1143.

5. Lee CT, Amaro RE. Exascale computing: a new dawn for computational biology. *Comput Sci Eng*. 2018;20(9):25.
6. Meyer T, D'Abramo M, Hospital A, et al. MoDEL (molecular dynamics extended library): a database of atomistic molecular dynamics trajectories. *Structure*. 2010;18:1399-1409.
7. Rysavy SJ, Beck DA, Daggett V. Dynameomics: data-driven methods and models for utilizing large-scale protein structure repositories for improving fragment-based loop prediction. *Protein Sci*. 2014;23:1584-1595.
8. Elofsson A, Hess B, Lindahl E, Onufriev A, van der Spoel D, Wallqvist A. Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. *PLoS Comput Biol*. 2019;15, 1:e1006649.
9. Wodak SJ, Paci E, Dokholyan NV, et al. Allostery in its many disguises: from theory to applications. *Structure*. 2019;27:566-578.
10. Amor BR, Schaub MT, Yaliraki SN, Barahona M. Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nat Commun*. 2016;7:1-13.
11. Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A. Protein contact networks: an emerging paradigm in chemistry. *Chem Rev*. 2013;113:1598-1613.
12. Yao XQ, Momin M, Hamelberg D. Establishing a framework of using residue-residue interactions in protein difference network analysis. *J Chem Inf Model*. 2019;59:3222-3228.
13. Serçinoğlu O, Ozbek P. GRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations. *Nucleic Acids Res*. 2018;46:W554-W562.
14. Nguyen HC, Zecchina R, Berg J. Inverse statistical problems: from the inverse Ising problem to data science. *Adv Phys*. 2017;66:197-261.
15. Wang E, Sun H, Wang J, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev*. 2019;119:9478-9508.
16. Di Nola A, Berendsen HJ, Edholm O. Free energy determination of polypeptide conformations generated by molecular dynamics. *Macromolecules*. 1984;17(10):2044-2050.
17. Singh S, Bowman GR. Quantifying allosteric communication via both concerted structural changes and conformational disorder with CARDS. *Journal of Chem Theory Comput*. 2017;13:1509-1517.
18. Wang S, Herzog ED, Kiss IZ, et al. Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks. *Proc Natl Acad Sci U S A*. 2018;115:9300-9305.
19. Bowerman S, Wereszczynski J. Detecting allosteric networks using molecular dynamics simulation. In: *Methods in Enzymology*. Vol 578. Academic Press Inc; 2016:429-447.
20. Hacısuleyman A, Erman B. Entropy transfer between residue pairs and Allostery in proteins: quantifying allosteric communication in ubiquitin. *PLoS Comput Biol*. 2017;13:e1005319.
21. Gasper PM, Fuglestad B, Komives EA, Markwick PR, McCammon JA. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci U S A*. 2012;109:21216-21222.
22. Melo MCR, Bernardi RC, de la Fuente-Nunez C, Luthey-Schulten Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J Chem Phys*. 2020;153:134104.
23. Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. *Proteins*. 2006;62:1053-1061.
24. DuBay KH, Bothma JP, Geissler PL. Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Comput Biol*. 2011;7(9):e1002168.
25. Marks DS, Colwell LJ, Sheridan R, et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS One*. 2011;6(12):e28766.
26. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J*. 2001;80(1):505-515.
27. Moritsugu K, Smith JC. Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance hessian. *Biophys J*. 2007;93:3460-3469.
28. Altis A, Otten M, Nguyen PH, Hegger R, Stock G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J Chem Phys*. 2008;128(4):245102.
29. Liu J, Amaral LAN, Keten S. Conformational stability of the bacterial adhesin, FimH, with an inactivating mutation. *Proteins*. 2021;89:276-288.
30. Pröpster JM, Yang F, Rabbani S, Ernst B, Allain FH-T, Schubert M. Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. *Proc Natl Acad Sci U S A*. 2016;113:E4170-E4179.
31. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367:1260-1263.
32. Woo H, Park SJ, Choi YK, et al. Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *Journal of Physical Chemistry B*. 2020;124:7128-7137.
33. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996;14:33-38.
34. Kalas V, Pinkner JS, Hannan TJ, et al. Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. *Sci Adv*. 2017;3:e1601944.
35. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26:1781-1802.
36. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14(1):71-73.
37. Altis A, Nguyen PH, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys*. 2007;126:244111.
38. Karami Y, Bitard-Feildel T, Laine E, Carbone A. "Infostery" analysis of short molecular dynamics simulations identifies highly sensitive residues and predicts deleterious mutations. *Sci Rep*. 2018;8:16126.
39. Sittel F, Jain A, Stock G. Principal component analysis of molecular dynamics: on the use of Cartesian vs. internal coordinates. *J Chem Phys*. 2014;141:014111.
40. Mendez R, Bastolla U. Torsional network model: Normal modes in torsion angle space better correlate with conformational changes in proteins. *Phys Rev Lett*. 2010;104:228103.
41. Trong IL, Aprikian P, Kidd BA, et al. Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like beta sheet twisting. *Cell*. 2010;141(4):645-655.
42. Sauer MM, Jakob RP, Eras J, et al. Catch-bond mechanism of the bacterial adhesin FimH. *Nat Commun*. 2016;7:10738.
43. Rodriguez VB, Kidd BA, Interlandi G, Tchesnokova V, Sokurenko EV, Thomas WE. Allosteric coupling in the bacterial adhesive protein FimH. *J Biol Chem*. 2013;288:24128-24139.
44. Rabbani S, Fiege B, Eris D, et al. Conformational switch of the bacterial adhesin FimH in the absence of the regulatory domain: engineering a minimalistic allosteric system. *J Biol Chem*. 2018;293:1835-1849.
45. Gonzalez-Gil A, Porell RN, Fernandes SM, et al. Sialylated keratan sulfate proteoglycans are Siglec-8 ligands in human airways. *Glycobiology*. 2018;28:786-801.
46. Henderson R, Edwards RJ, Mansouri K, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural and Molecular Biology*. 2020;27:925-933.
47. Cruz, M., Frederick, T., Singh, S., Vithani, N., Zimmerman, M., ... Bowman, G. Discovery of a cryptic allosteric site in Ebola's

- 'undruggable' VP35 protein using simulations and experiments, bioRxiv, 2020e.02.09.940510 2020.
48. Knoverek CR, Mallimadugula UL, Singh S, et al. Opening of a cryptic pocket in β -lactamase increases penicillinase activity. *Proc Natl Acad Sci U S A*. 2021;118:e2106473118.
 49. Grant BJ, Skjærven L, Yao X. The Bio3D packages for structural bioinformatics. *Protein Sci*. 2021;30:20-30.
 50. Keedy DA, Georgiev I, Triplett EB, Donald BR, Richardson DC, Richardson JS. The role of local backrub motions in evolved and designed mutations. *PLoS Comput Biol*. 2012;8:e1002629.
 51. Kisiela DI, Magala P, Interlandi G, et al. Toggle switch residues control allosteric transitions in bacterial adhesins by participating in a concerted repacking of the protein core. *PLoS Pathog*. 2021;17:e1009e440.
 52. Interlandi G, Thomas WE. Mechanism of allosteric propagation across a β -sheet structure investigated by molecular dynamics simulations. *Proteins*. 2016;84:990-1008.
 53. Freeman S, Birrell HC, D'Alessio K, Erickson-Miller C, Kikly K, Camilleri P. A comparative study of the asparagine-linked oligosaccharides on siglec-5, siglec-7 and siglec-8, expressed in a CHO cell line, and their contribution to ligand recognition. *Eur J Biochem*. 2001;268:1228-1237.
 54. Ke Z, Oton J, Qu K, et al. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*. 2020;588:498-502.
 55. Casalino L, Gaieb Z, Goldsmith JA, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci*. 2020;610:1722-1734.
 56. D.E. Shaw Research. Molecular dynamics simulations related to SARS-CoV-2. D. E. Shaw Research Technical Data; 2020. https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/
 57. Zimmerman MI, Porter JR, Ward MD, et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat Chem*. 2021;13:651-659.
 58. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature*. 2014;508:331-339.
 59. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007;450:913-916.
 60. De Domenico M, Solé-Ribalta A, Cozzo E, et al. Mathematical formulation of multilayer networks. *Phys Rev X*. 2014;3:041022.
 61. Peixoto TP. Bayesian stochastic blockmodeling. In: *Advances in Network Clustering and Blockmodeling*. Vol 12. John Wiley & Sons, Ltd; 2019:289-332.
 62. Banerjee A, Pathak J, Roy R, Restrepo JG, Ott E. Using machine learning to assess short term causal dependence and infer network links. *Chaos*. 2019;29:121104.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu J, Amaral LAN, Keten S. A new approach for extracting information from protein dynamics. *Proteins*. 2023;91(2):183-195. doi:[10.1002/prot.26421](https://doi.org/10.1002/prot.26421)