A Practical Covariance-Based Method for Efficient Detection of Protein-Protein Attractive and Repulsive Interactions in Molecular Dynamics Simulations

Mert Golcuk[1] and Mert Gur[1,*]

[1] Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, 15213, USA

* Corresponding Author, gurmert@pitt.edu

ABSTRACT

Molecular dynamics of large protein–protein complexes requires scalable analysis. We present a correlation-based workflow that systematically identifies stabilizing salt bridges, hydrogen bonds, hydrophobic contacts, and destabilizing repulsive interactions across extensive interfaces. By constructing an inter-protein covariance matrix, filtering residue pairs based on distance, and identifying interactions underlying these correlations, our method focuses computational resources on the most relevant regions of the interface while preserving a high level of detail.

MAIN

As both hardware and software technologies continuously advance, the size of protein systems and the length of all-atom molecular dynamics (MD) simulations that can be performed are expanding dramatically. A decade ago, typical simulations involved tens of thousands of atoms and spanned only a few tens of nanoseconds; today, it is not uncommon to simulate hundreds of thousands (or even millions[1-3]) of atoms over microsecond timescales. Consequently, whereas identifying interactions at protein–protein interfaces was once manageable through simple visual inspection or basic scripts, modern system sizes and simulation lengths demand more efficient methods to identify and characterize interactions across large protein–protein interfaces. Examples include interactions at oligomerization surfaces of homomers or heteromers, antibody–antigen complexes, motor proteins binding to microtubules, and microtubule-associated proteins (MAPs) binding to microtubules. For instance, the nanobody H11-H4 that binds the SARS-CoV-2 Spike protein has a binding interface spanning approximately 78 residues, while kinesin and dynein each have microtubule interaction surfaces comprising at least 101 residues.

Although a variety of publicly available codes and in-house tools exist to evaluate specific protein–protein interactions, they often rely on criteria such as distance and angle cutoffs for hydrogen bonds[4] or distance-based cutoffs for salt bridge determination[5]. However, such criteria-based methods can become computationally expensive and challenging to scale to larger systems and longer simulations, as the number of pairwise calculations required to identify interacting residues increases significantly with the size of the system and the duration of the simulation. Furthermore, identifying hydrophobic interactions through simple distance cutoffs is not straightforward because the presence of a water molecule can disrupt hydrophobic interaction. Additionally,

detecting destabilizing interactions such as unfavorable electrostatic contacts between like-charged residues is complex, since most criteria-based approaches focus primarily on stabilizing interactions and lack specific metrics to systematically identify or quantify these repulsive interactions.

Here, we introduce a new method that integrates and extends our existing approach[6], in which we demonstrated that constructing a covariance matrix from MD simulations can be used to efficiently identify both the presence and strength of interactions between two bound peptides from MD simulations. This results in a powerful, efficient, and user-friendly approach for identifying both stabilizing interactions (e.g., salt bridges, hydrogen bonds, hydrophobic interactions) and destabilizing interactions (e.g., repulsion between like-charged residues) from MD simulations. Our new methodology consists of three straightforward steps, demonstrated using an MD trajectory of a protein–protein complex. First, a covariance matrix is generated using the C_α atoms from the MD trajectory (Fig. 1A shows an example of an inter-protein covariance matrix computed between the nanobody H11-H4 and the SARS-CoV-2 Spike protein). Next, a distance cutoff is applied based on the average atomic coordinates used during the covariance matrix construction, setting all matrix elements beyond this cutoff to zero and thereby yielding a "close-contact" covariance matrix. This filtering step highlights residue pairs that are both spatially proximal and dynamically coupled, as illustrated in Fig. 1B, where a cutoff of 11 Å is applied for positively correlated pairs and 13 Å for negatively correlated pairs. The positive and negative values in this matrix indicate possible stabilizing (attractive) and destabilizing (repulsive) interactions, respectively, which will be evaluated in the next step. Focusing on the close-contact, correlated residue pairs identified by the filtering, the types and frequencies of specific interactions, including salt bridges, hydrogen bonds, hydrophobic contacts, and same-charge electrostatic repulsions, are then calculated across the simulation trajectory, significantly reducing the computational effort by restricting the analysis to residues most likely to interact. Fig.1C illustrates the resulting interaction frequency map between the nanobody H11-H4 and the SARS-CoV-2 Spike protein after these three steps, and Fig.1D visually depicts these interactions. By following this workflow, researchers can streamline protein–protein interaction analyses, reduce data complexity, and achieve clearer insights that can guide subsequent computational and experimental investigations.

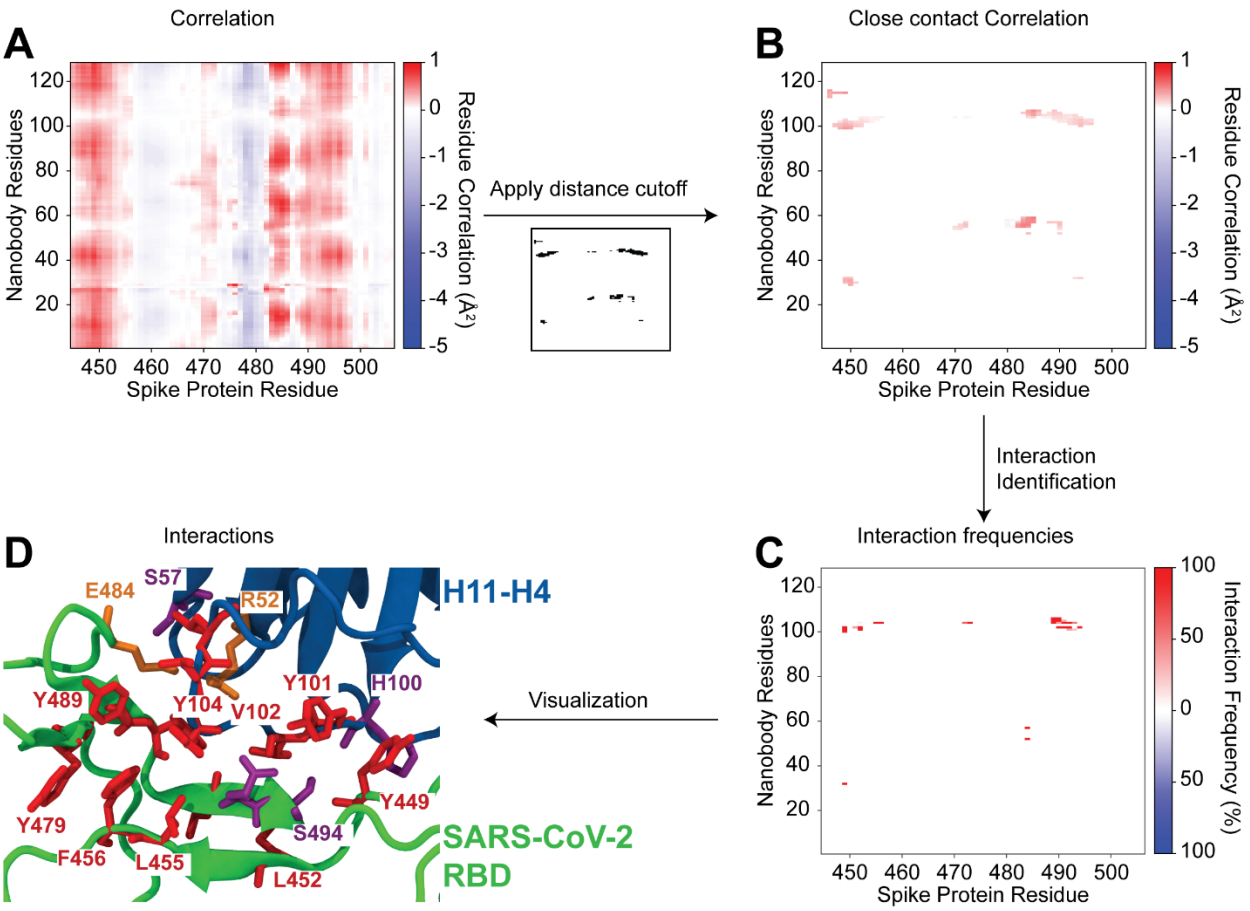


Figure 1 Methodology overview *(A) Inter-protein covariance matrix calculated using C_α atoms between nanobody H11-H4 residues and SARS-CoV-2 Spike protein residues. Red and blue colors indicate positive and negative residue correlations, respectively, suggesting attractive or repulsive interactions. (B) Close-contact correlation matrix obtained after applying a spatial distance cutoff (11 Å for positively correlated pairs, 13 Å for negatively correlated pairs), highlighting residue pairs that are spatially proximal and dynamically coupled. (C) Interaction frequency map displaying the percentage occurrence of specific interactions, including hydrogen bonds, salt bridges, hydrophobic contacts, and electrostatic repulsions, for residue pairs identified from the close-contact correlation matrix. (D) Structural visualization of identified interacting residues on nanobody H11-H4 (blue) and the SARS-CoV-2 Spike receptor-binding domain (green). Residues involved in stabilizing (attractive) and destabilizing (repulsive) interactions are explicitly labeled and colored by physicochemical properties: hydrophobic (red), positively charged (blue), negatively charged (orange), and polar (magenta), providing spatial context to interaction data.*

To further evaluate the accuracy and effectiveness of our methodology, we applied it to MD simulation trajectories (400 ns length each) from our previous studies[7, 8] investigating interactions between the nanobody H11-H4 and multiple SARS-CoV-2 Spike protein variants (Alpha, Beta, and Omicron). Our results demonstrate that our proposed method reliably and efficiently identifies key interactions, consistent with those found using conventional, more time-consuming interaction analysis workflows, in which all possible residue-residue interactions are exhaustively evaluated (Fig.2). In addition to simplifying interaction identification, our method offers a notable advantage: traditional approaches often treat interactions as binary (present or absent based on cutoffs) and generally do not reveal the spatial extent of an interaction's influence. In contrast, the correlation-based approach presented here captures not only the presence of interactions but also the dynamic coupling between residues. This enables insight into how specific attractive or repulsive interactions propagate through the protein interface and influence the local binding environment, an aspect that standard interaction analysis methods typically overlook.

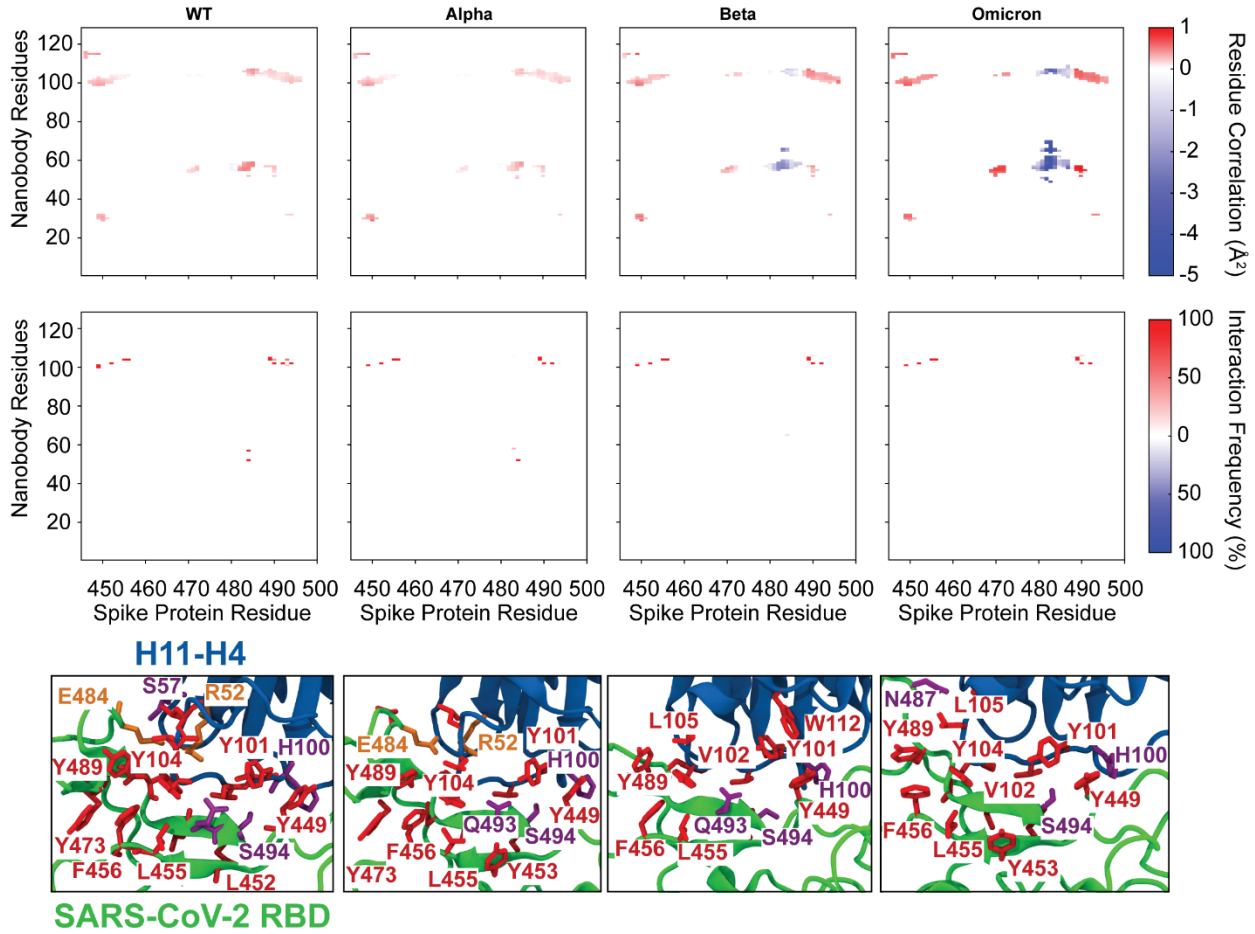


Figure 2 Comparison of nanobody H11-H4 interactions with wild-type and variant SARS-CoV-2 Spike proteins. *Top panels display close-contact covariance matrices illustrating residue correlations between nanobody H11-H4 and SARS-CoV-2 Spike receptor-binding domain (RBD) for WT, Alpha, Beta, and Omicron variants. Positive correlations (red) indicate attractive interactions, whereas negative correlations (blue) represent repulsive interactions. Middle panels show interaction frequency maps corresponding to each variant, indicating the percentage occurrence of specific residue-residue interactions. Bottom panels visualize interactions on protein structures, highlighting residues involved in stabilizing and destabilizing interactions. Residues are explicitly labeled and colored by physicochemical properties: hydrophobic (red), positively charged (blue), negatively charged (orange), and polar (magenta). Differences in interaction patterns between variants illustrate the impact of Spike protein mutations on nanobody binding.*

METHODS

Algorithm

The details and implementation of this methodology, comprising three main steps, are described as follows:

1. **Constructing the Inter-Protein Covariance Matrix**
   For a protein–protein complex, the inter-protein covariance matrix ($\mathbf{C}$) is computed as:

$$\mathbf{C} = \langle \left(\mathbf{R}_{protein1} - \langle \mathbf{R}_{protein1} \rangle\right)\left(\mathbf{R}_{protein2} - \langle \mathbf{R}_{protein2} \rangle\right)^{\mathrm{T}} \rangle \qquad \text{(Eq.1)}$$

   Here, *protein1* and *protein2* refer to the two interacting proteins within the complex. **R** denotes the configuration vector of a given conformation from an MD simulation, and $\langle \mathbf{R} \rangle$ represents the trajectory-averaged configuration vector. We adopt a coarse-grained model based on the C_α atom representation. Here, *n* and *m* represent the number of residues considered in *protein1* and *protein2* (either all C_α of the full sequences or selected subsets), respectively; resulting in $\mathbf{R}_{protein1}$ and $\mathbf{R}_{protein2}$ having dimensions of $3n$ and $3m$. However, the model can also be extended to an all-atom representation or to another atom selection, such as backbone atoms.

2. **Applying a Distance Cutoff to Identify Close Contact Correlations**
   As a next step, a spatial cutoff is applied to identify residue pairs that come within a defined "interaction distance." For simplicity, the trajectory-averaged coordinates ($\langle \mathbf{R}_{protein1} \rangle$ and $\langle \mathbf{R}_{protein2} \rangle$), already computed in the previous step, are used. Any elements of the covariance matrix corresponding to residue pairs whose average inter-residue distance falls outside the cutoff are set to zero and a new 'close-contact covariance matrix' is constructed.
   Conceptually, when two residues interact, they tend to maintain a preferred equilibrium distance, moving in concert: if one residue moves in a positive direction, the other typically moves similarly, and vice versa. In the covariance matrix, such a concerted motion between residues *i* and *j* leads to a positive ensemble average (i.e., a positive element (*i,j*) of the covariance matrix, $\langle \Delta\mathbf{R}i \cdot \Delta\mathbf{R}j \rangle$). If there is no interaction, the residues move independently, resulting in near-zero $\langle \Delta\mathbf{R}i \cdot \Delta\mathbf{R}j \rangle$. Conversely, if two residues strongly repel each other, they tend to move in opposite directions, resulting in a negative $\langle \Delta\mathbf{R}i \cdot \Delta\mathbf{R}j \rangle$.

3. **Performing Interaction Analysis**
   In this step, either 3A or 3B, or preferably both, can be performed to obtain more detailed insights into the protein–protein interactions.

   3A. Focused Interaction Calculations
   For each close-contact, correlated residue pair identified in the previous step, we assessed the types and frequencies of specific interactions as follows. Salt bridges were identified

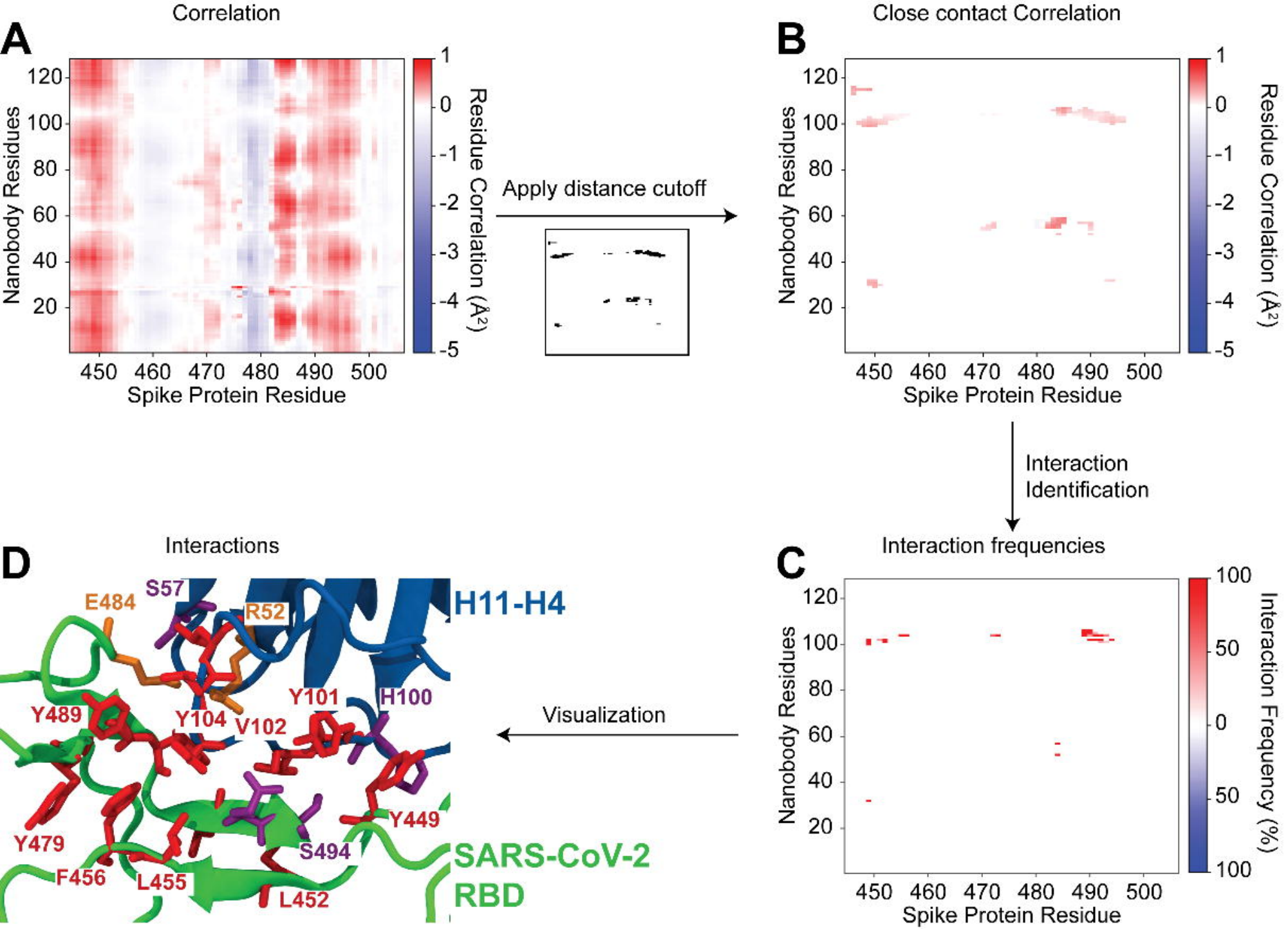
as interactions between basic nitrogen and acidic oxygen atoms within 6 Å[5]. Hydrogen bonds were defined as those in which the donor–acceptor distance was ≤3.5 Å and the donor–hydrogen–acceptor angle was ≤30° [30]. Hydrophobic interactions were characterized by contacts between side-chain carbon atoms within an 8 Å cutoff. Finally, repulsive correlations were determined when pairs of basic nitrogen or acidic oxygen atoms were within 12 Å of each other.

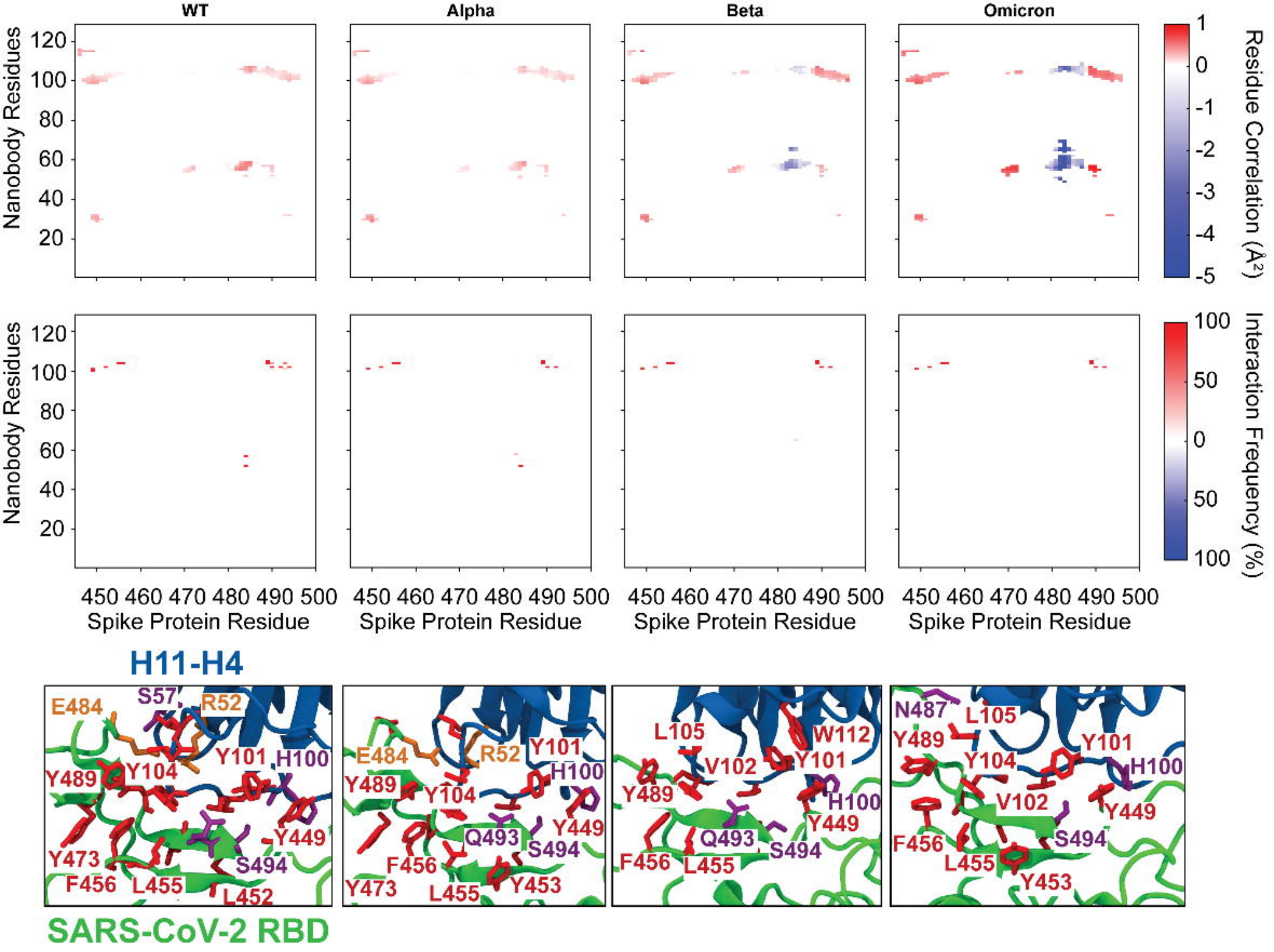
3B. Interaction Analysis through Visualization

The correlated, close-contact residue pairs identified in Step 3 are visualized on the protein structures using a molecular visualization program (e.g., VMD[9]) to identify those involved in either attractive or repulsive interactions. **Fig. 1D** shows the positions of the correlated close-contact residues displayed on the structures of the nanobody H11-H4 and the SARS-CoV-2 Spike protein.

REFERENCES

1. Can, S., Lacey, S., Gur, M., Carter, A.P. & Yildiz, A. Directionality of dynein is controlled by the angle and length of its stalk. *Nature* **566**, 407-410 (2019).
2. Gur, M. et al. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *J Chem Phys* **153**, 075101 (2020).
3. Golcuk, M., Yilmaz, S.Z., Yildiz, A. & Gur, M. The mechanism and energetics of the dynein priming stroke. *Structure* **32**, 603-610.e604 (2024).
4. Durrant, J.D. & McCammon, J.A. HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J Mol Graph Model* **31**, 5-9 (2011).
5. Beckstein, O., Denning, E.J., Perilla, J.R. & Woolf, T.B. Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open<-->closed transitions. *J Mol Biol* **394**, 160-176 (2009).
6. Gur, M. & Erman, B. Quasi-harmonic fluctuations of two bound peptides. *Proteins* **80**, 2769-2779 (2012).
7. Golcuk, M., Hacisuleyman, A., Erman, B., Yildiz, A. & Gur, M. Binding Mechanism of Neutralizing Nanobodies Targeting SARS-CoV-2 Spike Glycoprotein. *J Chem Inf Model* **61**, 5152-5160 (2021).
8. Golcuk, M. et al. SARS-CoV-2 Delta Variant Decreases Nanobody Binding and ACE2 Blocking Effectivity. *J Chem Inf Model* **62**, 2490-2498 (2022).
9. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).


A
Correlation
Nanobody Residues
Spike Protein Residue
Residue Correlation (Å²)
Apply distance cutoff
B
Close contact Correlation
Nanobody Residues
Spike Protein Residue
Residue Correlation (Å²)
Interaction Identification
C
Interaction frequencies
Nanobody Residues
Spike Protein Residue
Interaction Frequency (%)
Visualization
D
Interactions
H11-H4
SARS-CoV-2 RBD
S57
E484
R52
Y489
Y104
V102
Y101
H100
Y449
Y479
S494
F456
L455
L452


WT
Alpha
Beta
Omicron
Nanobody Residues
120
100
80
60
40
20
Residue Correlation (Å²)
1
0
-1
-2
-3
-4
-5
Interaction Frequency (%)
100
50
0
50
100
450 460 470 480 490 500
Spike Protein Residue
H11-H4
E484
S57
R52
Y101
H100
Y489
Y104
Y449
Y473
S494
F456
L455
L452
Q493
Y453
L105
W112
V102
N487
SARS-CoV-2 RBD