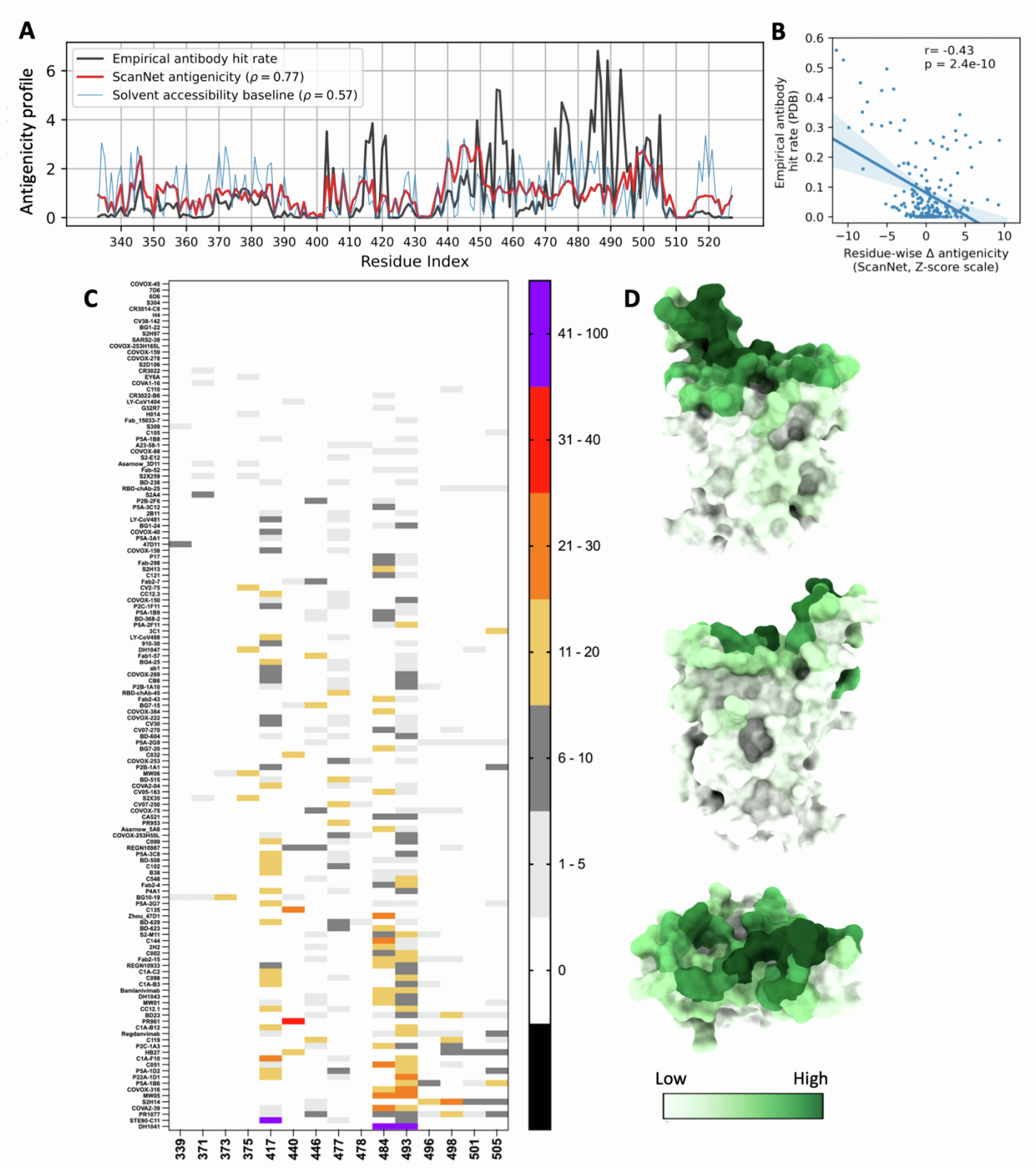
Cell Reports, Volume *41*

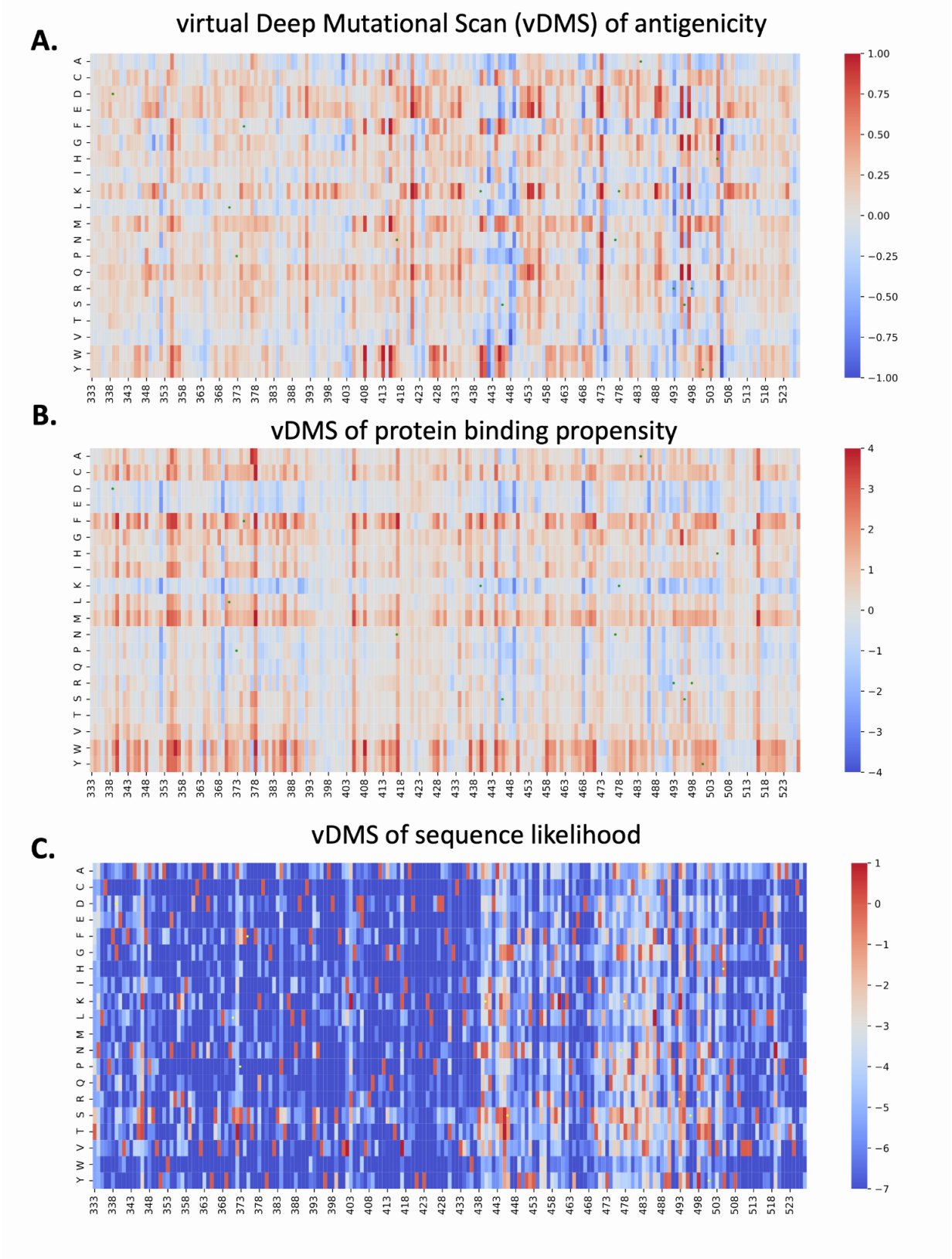
Supplemental information

Reduced B cell antigenicity of Omicron lowers host serologic response

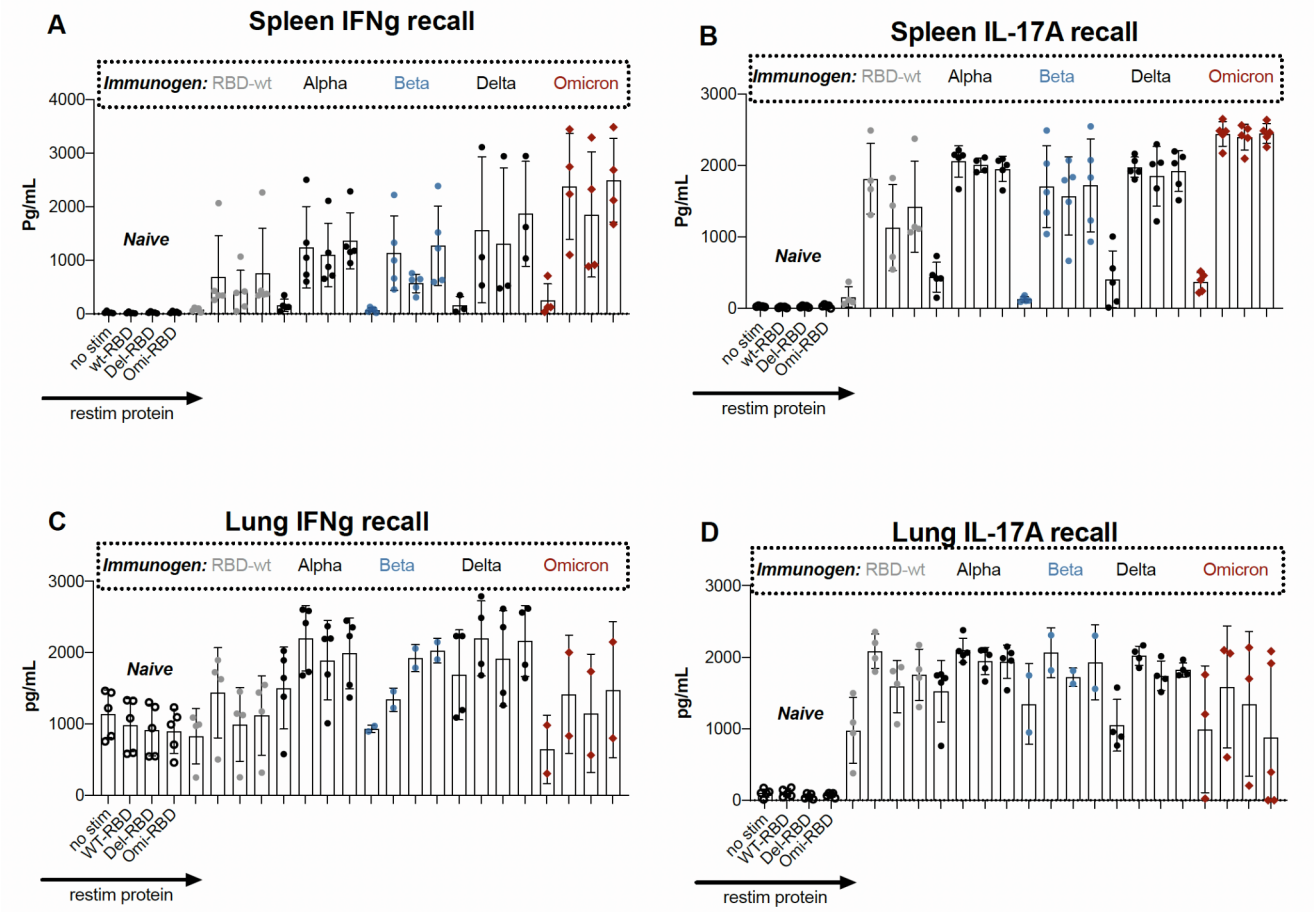
Jérôme Tubiana, Yufei Xiang, Li Fan, Haim J. Wolfson, Kong Chen, Dina Schneidman-Duhovny, and Yi Shi



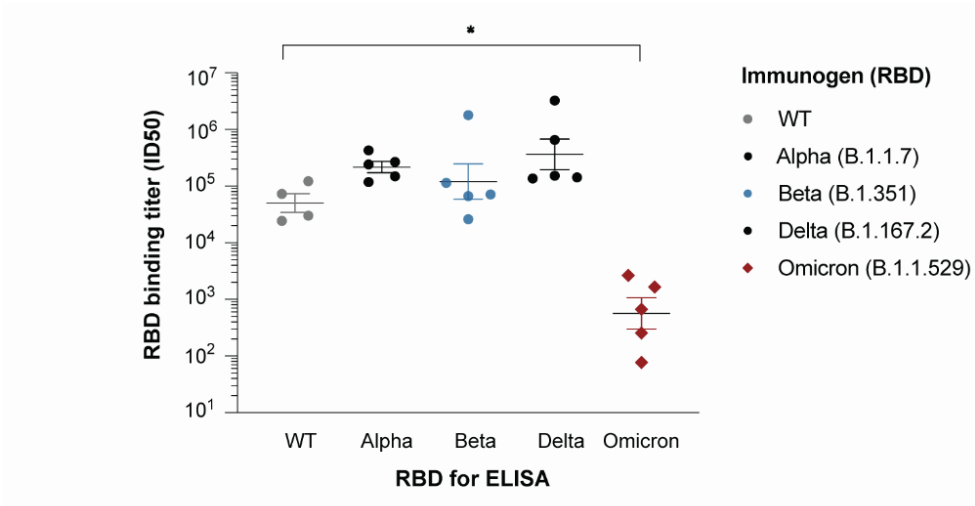
Supplementary Figure S1 (Related to Figure 1): Epitope distribution based on solved antibody-RBD structures and link with the ScanNet antigenicity profile. A ScanNet antigenicity profile for WT RBD (red) vs. the empirical antibody hit rate calculated based on PDB structures (black). RBD residue solvent accessibility is shown as a baseline (blue). The three curves were normalized to the mean value of 1 to facilitate the comparison. The solvent accessibility was computed within the spike trimer in the open state configuration (PDB identifier: 7e5r:A). **B** Scatter plot of the residue-wise empirical antibody hit rate (scaling: fraction of antibodies that bind the residue) against the change in ScanNet antigenicity of the Omicron variant (Z-score scale, using the residue-wise WT antigenicity distribution as reference, corresponds to the marker size of points in Figure 1B). Solid line and shaded area indicate respectively the linear regression fit and its 95% confidence interval. P-value calculated using Student t-test. **C** Omicron mutations (x-axis) vs. antibody binding. The color-coding corresponds to the fraction of the RBD residue atoms interacting with the antibody. **D** RBD structure (PDB: 7jvb) is colored according to the antibody hit rate (number of antibodies that interact with the residue).



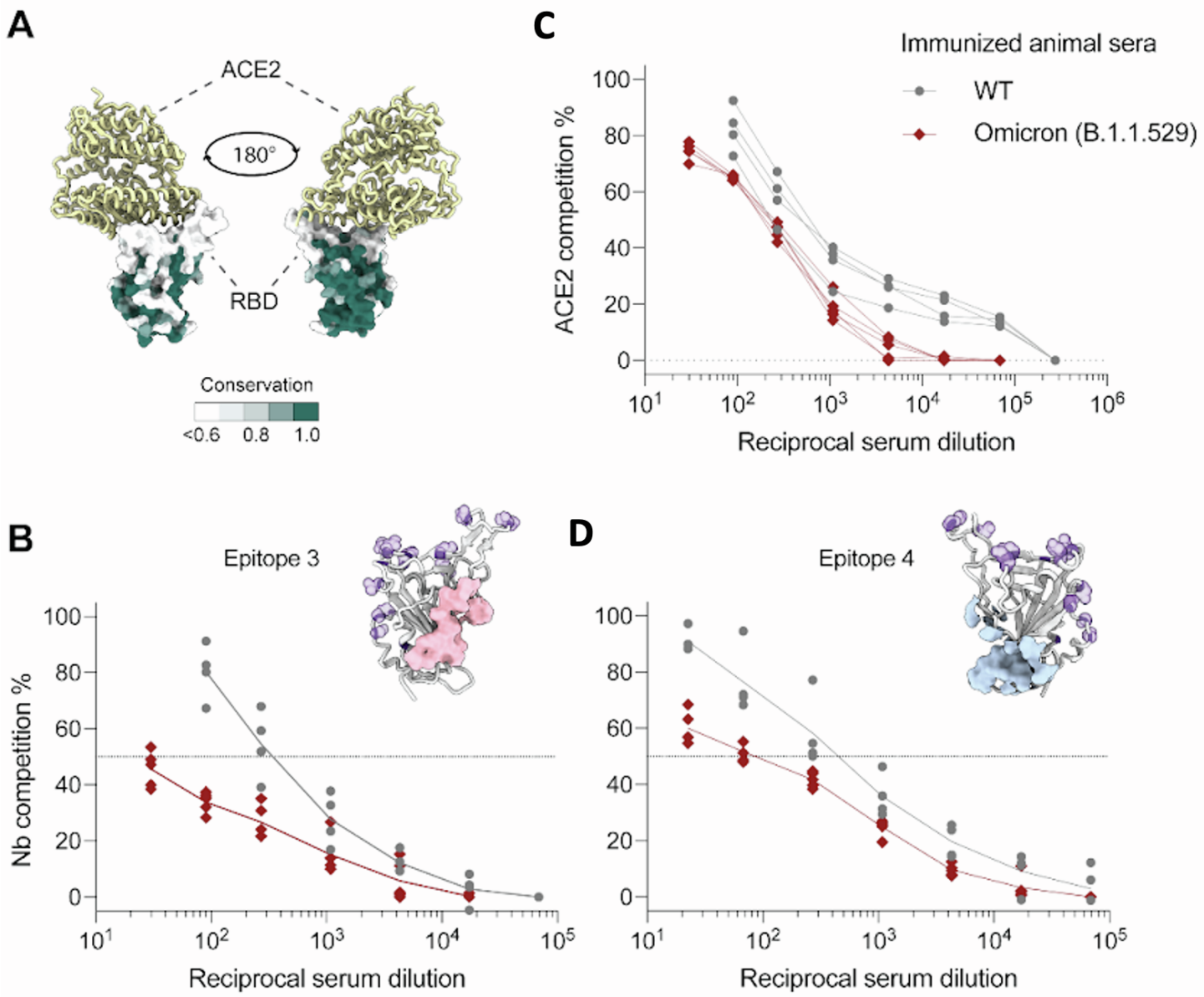
Supplementary Figure S2 (Related to Figure 1): virtual Deep Mutational Scans (vDMS) of WT RBD for antigenicity, protein binding propensity and sequence likelihood. A vDMS of antigenicity score predicted by ScanNet, see Methods. The difference between the *summed* antigenicity profiles (*over the whole domain*) of mutants and WT is shown. The corresponding distribution of entries is shown in **Figure 1F, Figure S8B**. Green dots indicate Omicron mutations. **B.** vDMS of protein binding propensity score predicted by ScanNet, see Methods. The difference between the *summed* protein binding propensity profiles *(over the whole domain)* of mutants and WT is shown. The corresponding distribution of entries is shown in **Supplementary Figure S6D**. Green dots indicate Omicron mutations. **C.** vDMS of sequence likelihood based on evolutionary records. The difference between the log-likelihood log P(S) of mutant sequence and wild type sequence is shown. The likelihood function was obtained by training a sequence generative model (Restricted Boltzmann Machine), on a multiple sequence alignment of betacoronaviruses RBDs (**Methods**). The vDMS were averaged over five trainings, each using different random seeds. The correlation to the expression level DMS performed in(Starr *et al.*, 2020) is shown in **Supplementary Figure S6D**. Yellow dots indicate Omicron mutations.



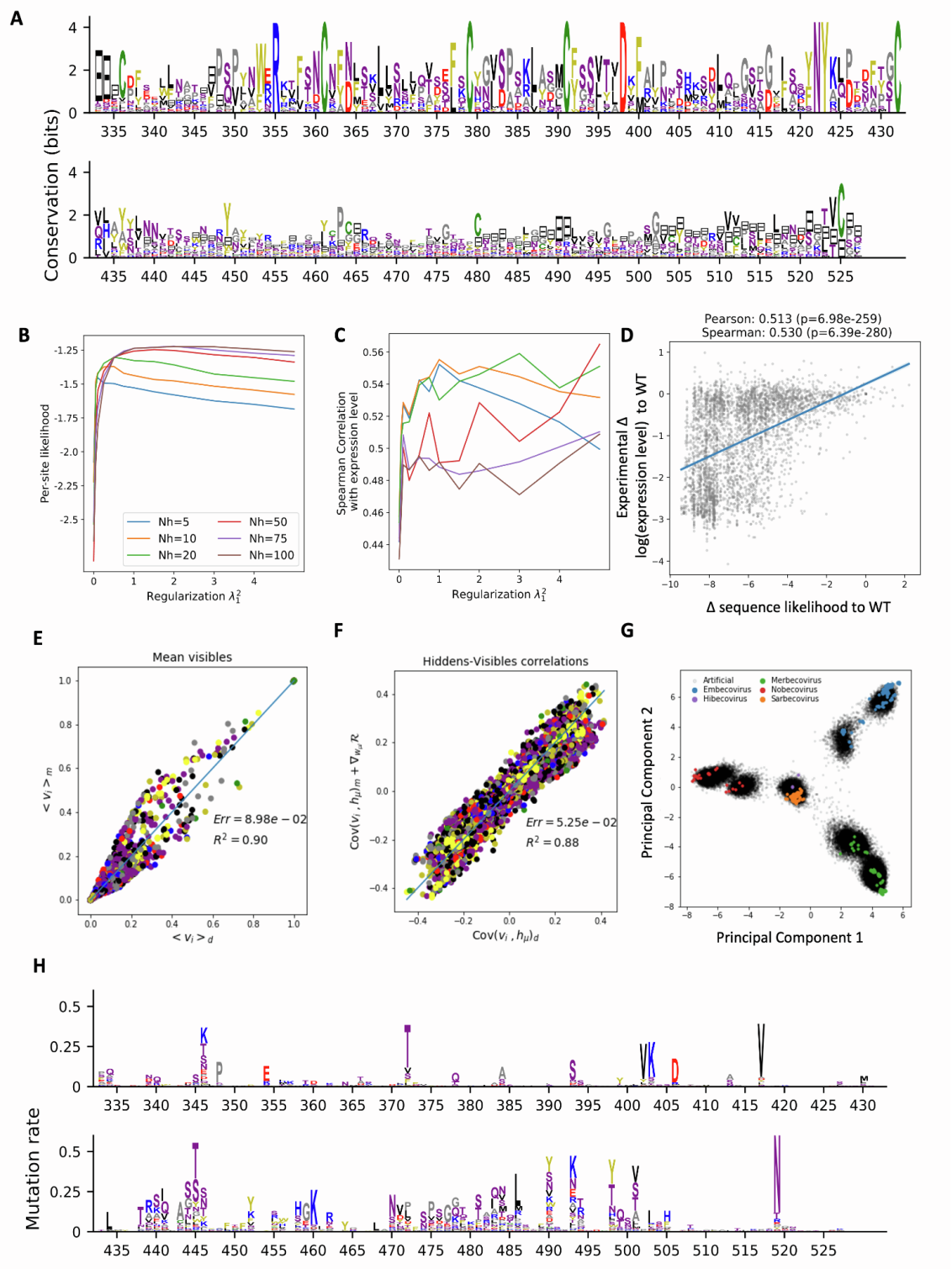
Supplementary Figure S3 (Related to Figure 2): The T cell recall responses in the RBD immunized mice. Splenocytes (**A, B**) or lung mononuclear cells (**C, D**) from naïve or RBD immunized mice were left unstimulated or stimulated with WT, Delta, or Omicron RBD for 72h. Culture supernatants were harvested for IFNg and IL-17 measurements by ELISA. Each data point represents an animal. Bars and whiskers represent mean +/- SD. Note: Some data points are missing due to technical difficulties in the mononuclear cell isolation from the inflamed tissues.



Supplementary Figure S4 (Related to Figure 2): ELISA of RBD-immunized mouse sera on day 25 against the corresponding antigen. Binding titer was calculated as the ID_{50} (reciprocal serum dilution that inhibits the 50% maximal RBD binding).

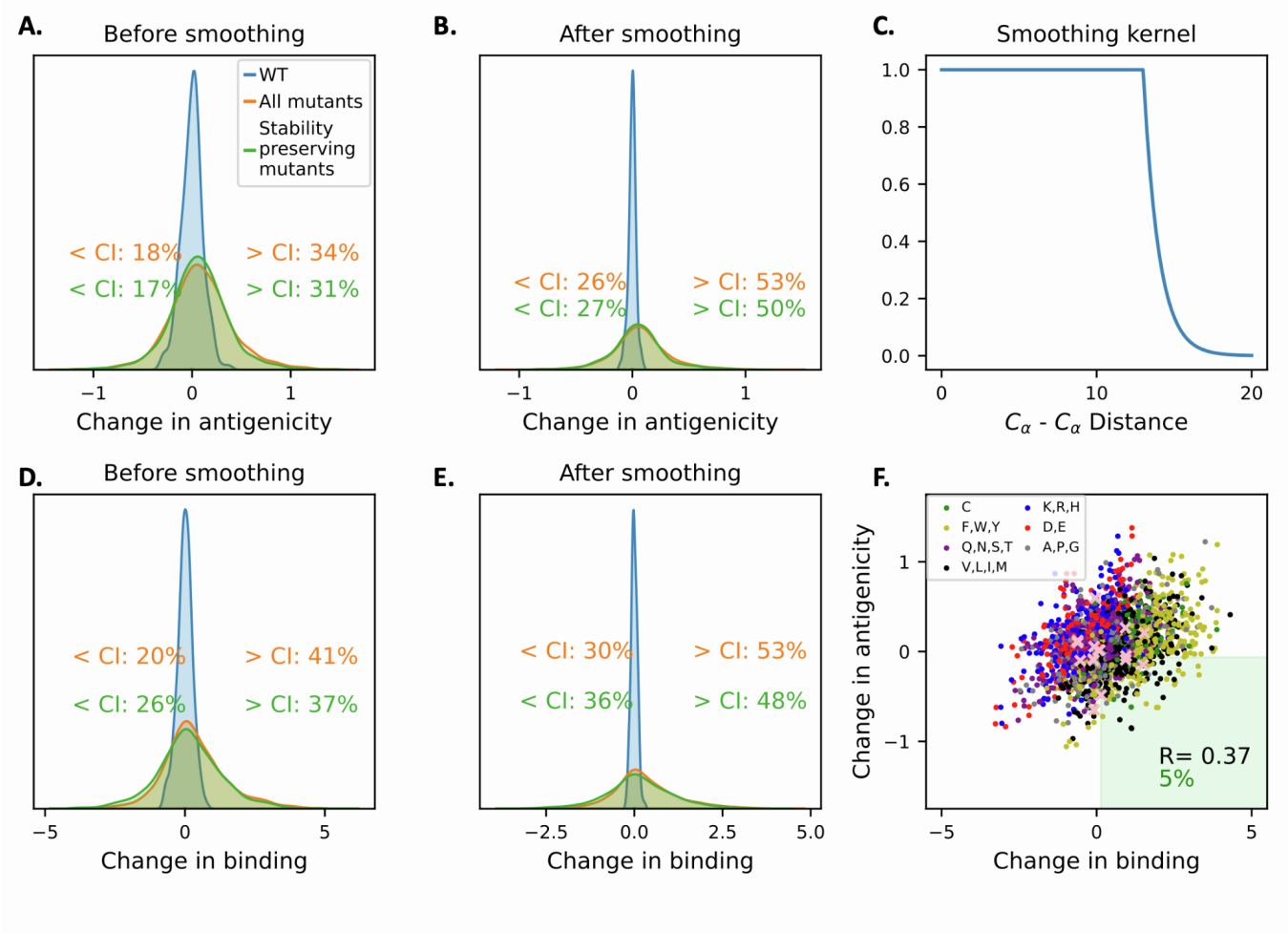


Supplementary Figure S5 (Related to Figure 1): Competition ELISA. **A**. Structural representations of the RBD-ACE2 complex. The sequence conservation of sarbecovirus RBD was presented in a color gradient, where 1.0 (in dark green) indicates that the residue is 100% conserved within all the sarbecovirus clades. **B-D**. Competitive ELISA between mice sera and (**C**) hACE2, (**B**) a high-affinity nanobody that targets a conserved RBD epitope (residues 337, 351-358, 396, 464, 466-468), or (**D**) a high-affinity nanobody that targets another conserved RBD epitope (residues 380, 381, 386, 390, 393, 428-431, 464, 514-522) for RBD binding. Each plot shows the percentage of remaining ACE2 or Nbs on the immobilized RBD in the presence of sera, expressed as reciprocal serum dilution. RBD was shown as gray ribbons. Mutated residues on Omicron were shown in purple. Distinct, conserved nanobody epitopes (3 and 4) were shown in pink and blue, respectively.



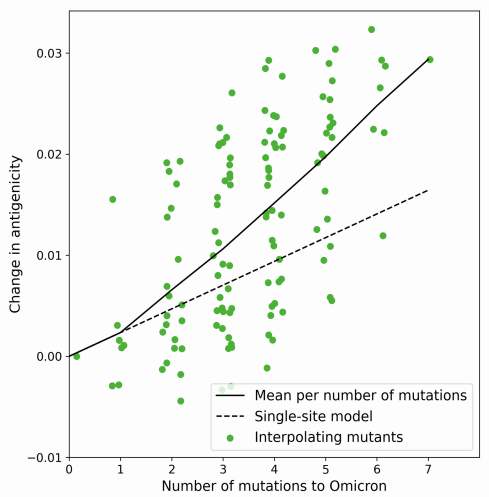
Supplementary Figure S6 (Related to STAR Methods): Training, validation and sampling of a sequence generative model for the RBD.

A. Sequence profile of the MSA of betacoronavirus RBDs identified in UniRef30. **B,C.** Hyperparameter search by cross-validation. (**B**) Cross-validation likelihood (divided by the number of sites) and (**C**) Spearman correlation between the change in sequence log-likelihood and change in expression around WT, as a function of the number of hidden units M and the regularization strength λ_1^2 of the RBM. The pseudo-likelihood (not shown) was also monitored and was highly correlated to the likelihood. Since the three metrics did not peak at the same position, we manually selected $M = 20, \lambda_1^2 = 0.5$ as a compromise. **D.** Scatter plot of the change in expression level (experimentally determined in(Starr *et al.*, 2020)) and the change in log-likelihood for all single point mutants (corresponding matrix shown in **Figure S3C**). **E-G.** Evaluation of the generative properties of the selected model. The distribution of samples generated by the model matched reasonably well the (**E**) single site frequency, (**F**) covariance between visible and hidden units and (**G**) overall topology of the distribution of natural sequences. **H.** Distribution of mutations to WT found in the set of 1000 artificial variants obtained by sampling from $P(S|D_{Hamming}(S, WT) = 15, N_{gaps} = 0)$. For each site, the height of each letter is proportional to the frequency of the corresponding amino acid in the generated set. The total height is proportional to the mutation rate of the site. As expected, RBS is the most variable region.



Supplementary Figure S7 (Related to STAR Methods): Noise estimation and reduction methodology for assessing the impact of all single point mutations on antigenicity and protein binding propensity.

A-B. Histogram of the total change in antigenicity concerning WT across all single point mutations before (**A**) and after (**B**) smoothing. Blue histogram represents 195 repeated runs of the WT sequence through the comparative modeling pipeline; it corresponds to the noise level induced by homology modeling. Text indicates the fraction of mutations outside of the [5%,95%] confidence interval. **C.** The smoothing kernel used for weighting residues in the neighborhood of the mutation. **D-E.** Histogram of the total change in protein binding propensity with respect to WT across all single point mutations before (D) and after (E) smoothing. **F.** Change in protein binding propensity vs. change in antigenicity for all single-point mutants. Each point corresponds to a single mutation, colored by the type of amino acid (green: cysteine, gold: aromatic, purple: polar, black: hydrophobic, blue: positively charged, red: negatively charged, gray: tiny/proline). The green shaded region denotes mutations that simultaneously increase protein binding propensity and decrease antigenicity; they form a small since both properties are correlated. Pink crosses indicate Omicron mutations.



Supplementary Figure S8 (Related to STAR Methods): Epistatic effects for antigenicity. Starting from the Omicron sequence, all 2^7 combinations of the seven antigenicity-increasing mutations that reverse to WT sequence are modeled and their RBS antigenicity is calculated. The graph shows the change in antigenicity as a function of mutation number. Solid and dashed lines indicate average and predicted average from single mutations. On average, the combined effect of mutations is larger than the sum of the individual effects.