nature biomedical engineering

Article

https://doi.org/10.1038/s41551-023-01074-6

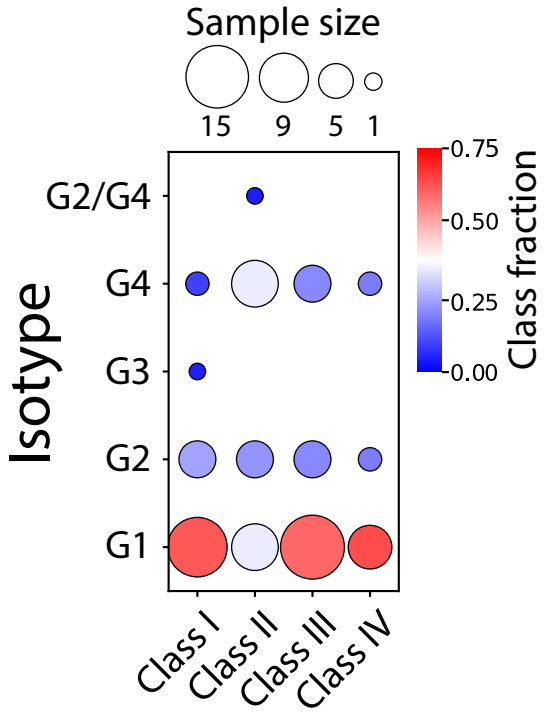
Optimization of therapeutic antibodies for reduced self-association and non-specific binding via interpretable machine learning

In the format provided by the authors and unedited

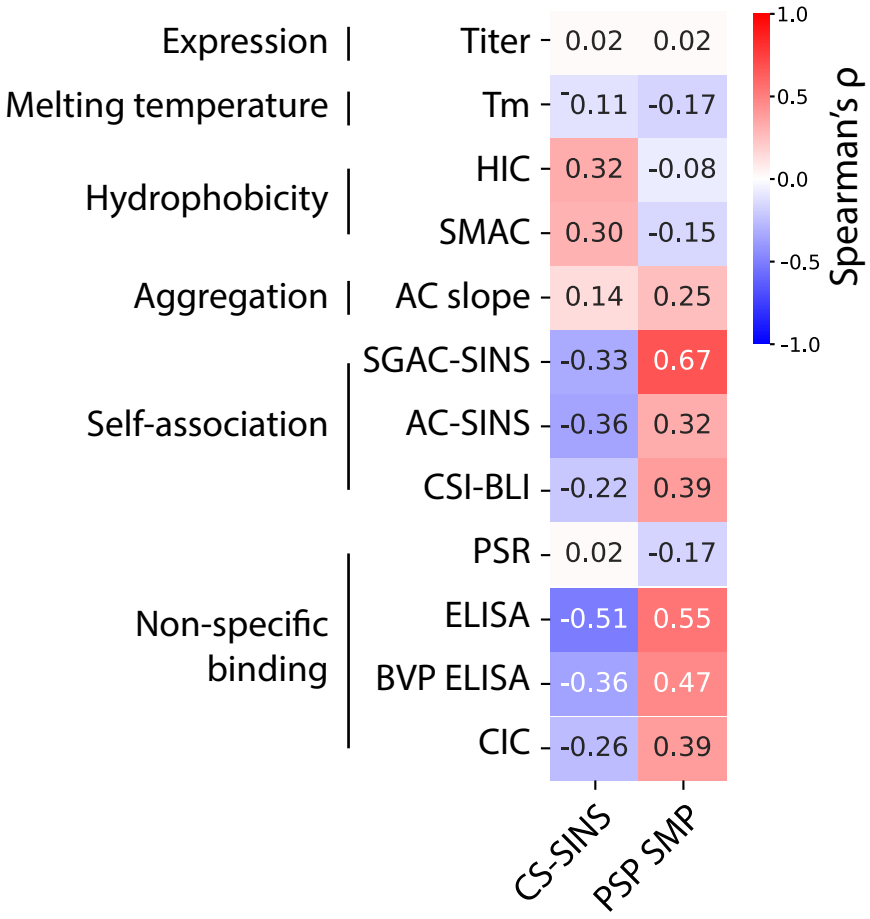
Supplementary Table 1 | Summary of the designed mutations. The VH (top table) and VL (bottom table) mutations were designed based on several criteria, including the frequency of the wild-type residue and potential mutations in human repertoires, as well as the rank of the most frequent residues at each site in the Fv region in human repertoires (as judged by the abYsis database). The germline residue and location of each site were also considered. The color codes are red for negative charge, blue for positive charge, purple for mutations that were different than germline, and green for mutations to germline residues. FR is framework and freq. is frequency.

	Mutation position	WT	Mutation	WT freq. (%)	Mutation freq.	Mutation freq. rank	Germline residue	Location
cinpanemab	H13	E	K	1	52		K	FR
	H43	Q	K	21	70		K	FR
	H64	E	K	2	63		K	CDR2
	H73	D	N	5	40		D	FR
gantenerumab	H100G	R	F	3	12		N/A	CDR3
panitumumab	H62	S	K	73	16	2nd	S	CDR2
	H82A	S	R	31	9	3rd	S	FR
	H95	D	V	23	7	4th	D	CDR3

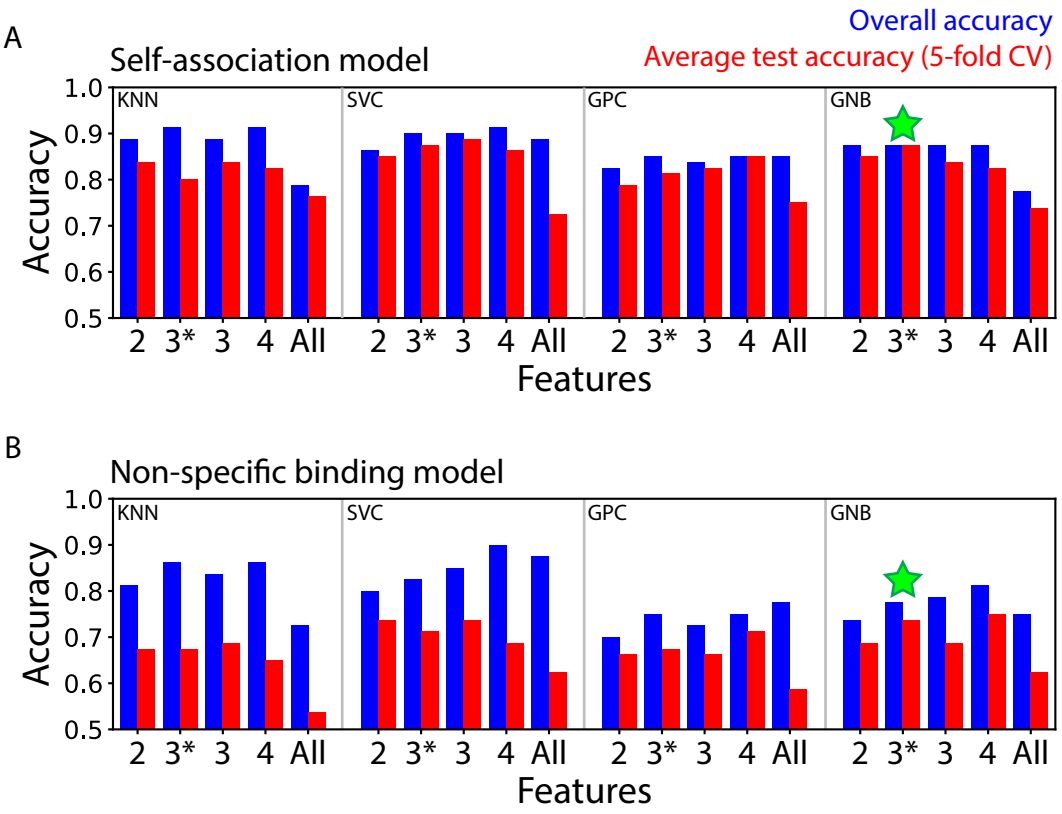
	Mutation position	WT	Mutation	WT Nat. Div.	Mutation Nat. Div.	Mutation Nat. Div. Rank	Germline residue	Location
cinpanemab	L3	E	V	11	56		E	FR
	L26	E	S	<1	75		D	CDR1
	L60	E	S	9	24		E	FR
gantenerumab	L18	R	T	54	16	2nd	R	FR
	L24	R	T	44	15	2nd	R	CDR1
	L24	R	S	44	24	3rd	R	CDR1
	L45	R	V	26	9	3rd	L	FR
panitumumab	L24	Q	R	26	9	3rd	Q	CDR1
	L70	D	T	41	18	3rd	D	FR
	L77	S	R	33	19	3rd	S	FR



Supplementary Fig. 1 | Analysis of the original isotypes of the 80 clinical-stage antibodies evaluated in this study. The frequency of original antibody isotype occurring in each identied class of behavior (TABS database).



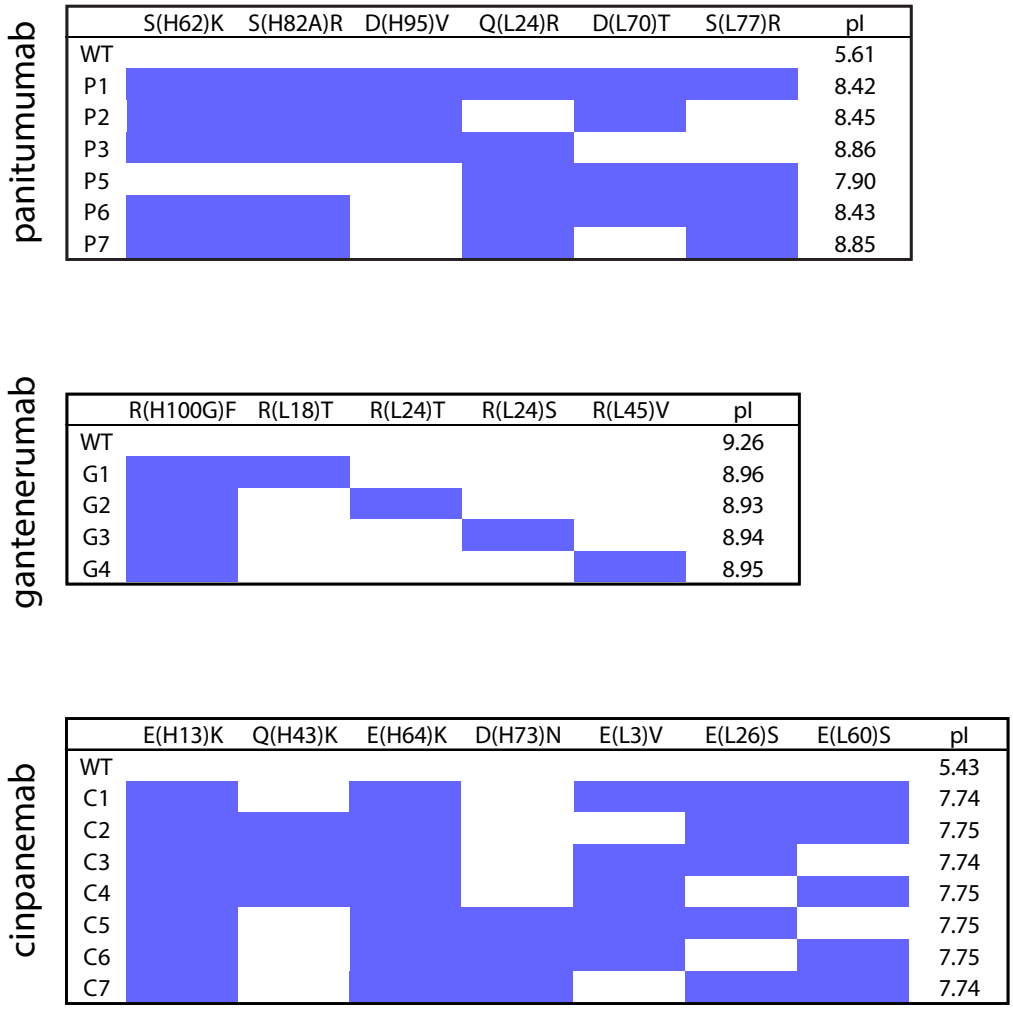
Supplementary Fig. 2 | Correlation analysis between the self-association and non-specific binding measurements in this study with other biophysical measurements reported previously. Spearman's ρ values are reported a common set of 50 clinical-stage antibodies evaluated in this study relative to those in a previous study (Jain et al., PNAS, 2017). The previous work only measured self-association in a physiological solution (pH 7.4, PBS), which results in poor correlations with the CS-SINS measurements performed using a standard formulation condition (pH 6, 10 mM histidine). The previously reported assays correspond to measurements of antibody expression (titer), melting temperature (Tm), self-association (SGAC-SINS, AC-SINS, CSI-BLI), non-specific binding (PSR, ELISA, BVP ELISA, CIC), aggregation (AC slope), and hydrophobicity (HIC, SMAC).



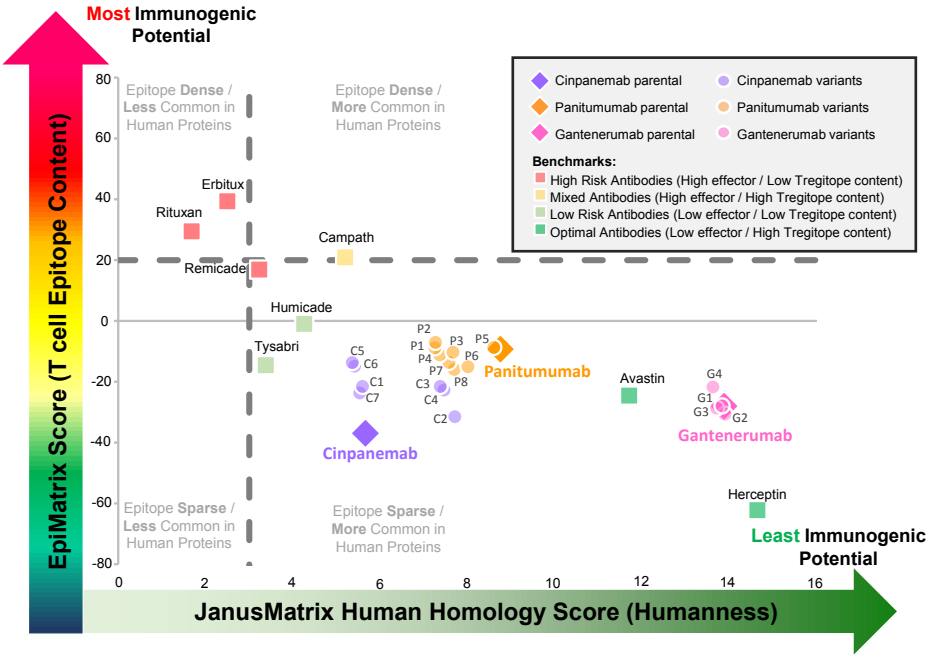
C

	Self-association model	Non-specific binding model
Overall accuracy	88%	78%
Balanced accuracy	88%	82%
Precision	82%	56%
Recall	88%	91%
F1	85%	69%

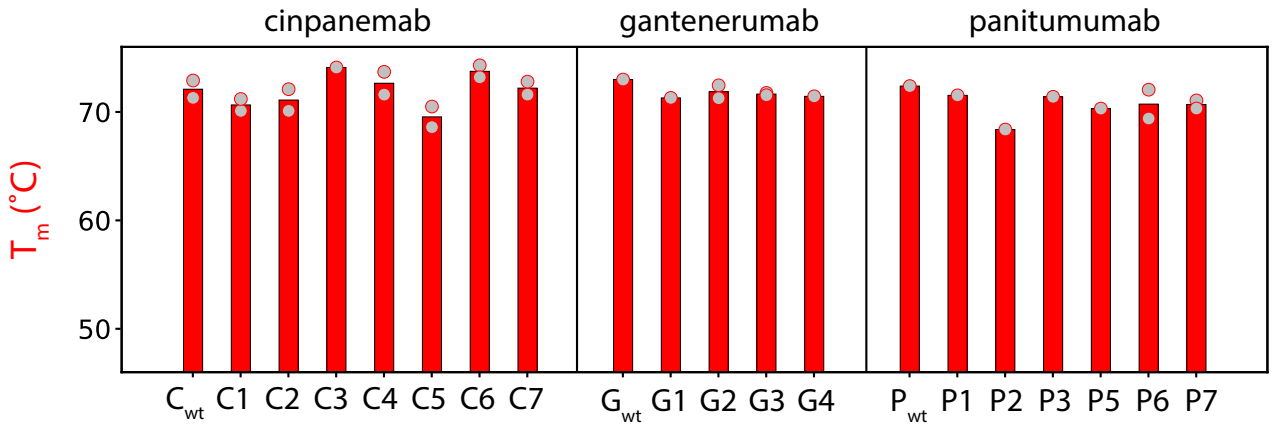
Supplementary Fig. 3 | Performance of four interpretable machine learning model architectures with differ-ent numbers of features. (A-B) Classifier models [K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Gaussian Process Classifier (GPC), and Gaussian naïve Bayes (GNB)] were benchmarked using average test accuracy for fivefold cross validation and overall model accuracy as performance metrics. Five combinations of feature numbers and types (e.g., ratios) were evaluated, namely, "2", "3*", "3", "4", and "all". Feature sets "2", "3", "4", and "all" were trained using two, three, four, and thirty-three (all) features, respectively. In the "3*" feature set, one independent feature and ratio of two independent features was used. For the (A) self-association and (B) non-specific binding models, Gaussian naïve Bayes classifier models using "3*" features (green star) generally display the best overall performance. (C) Summary of the performance of the final Gaussian naïve Bayes classifier models.



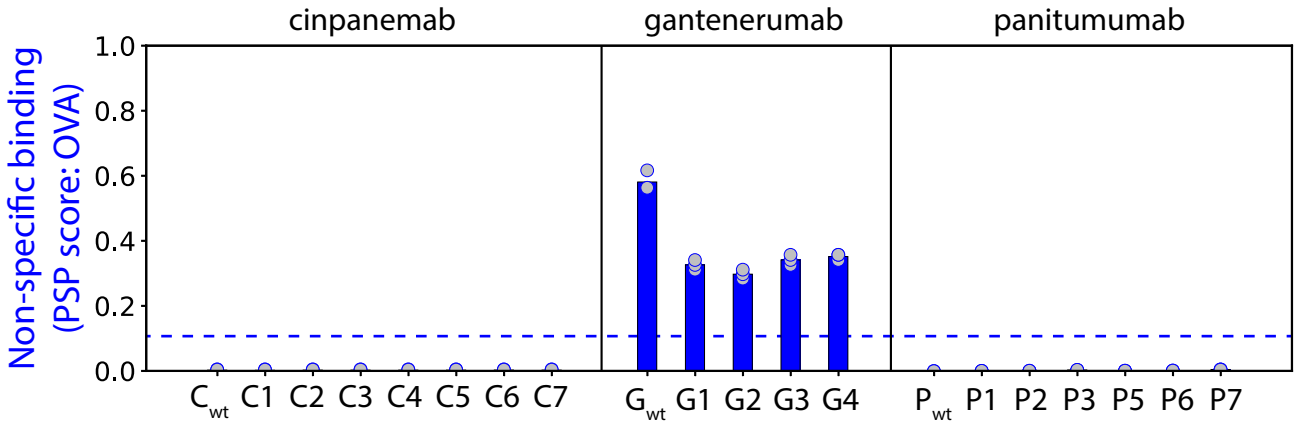
Supplementary Fig. 4 | Summary of designed mutations in the heavy and light chains. Combinations of designed mutations predicted to reduce non-affinity interactions while maintaining affinity interac-tions. Structure-based isoelectric points (pIs) are also reported.



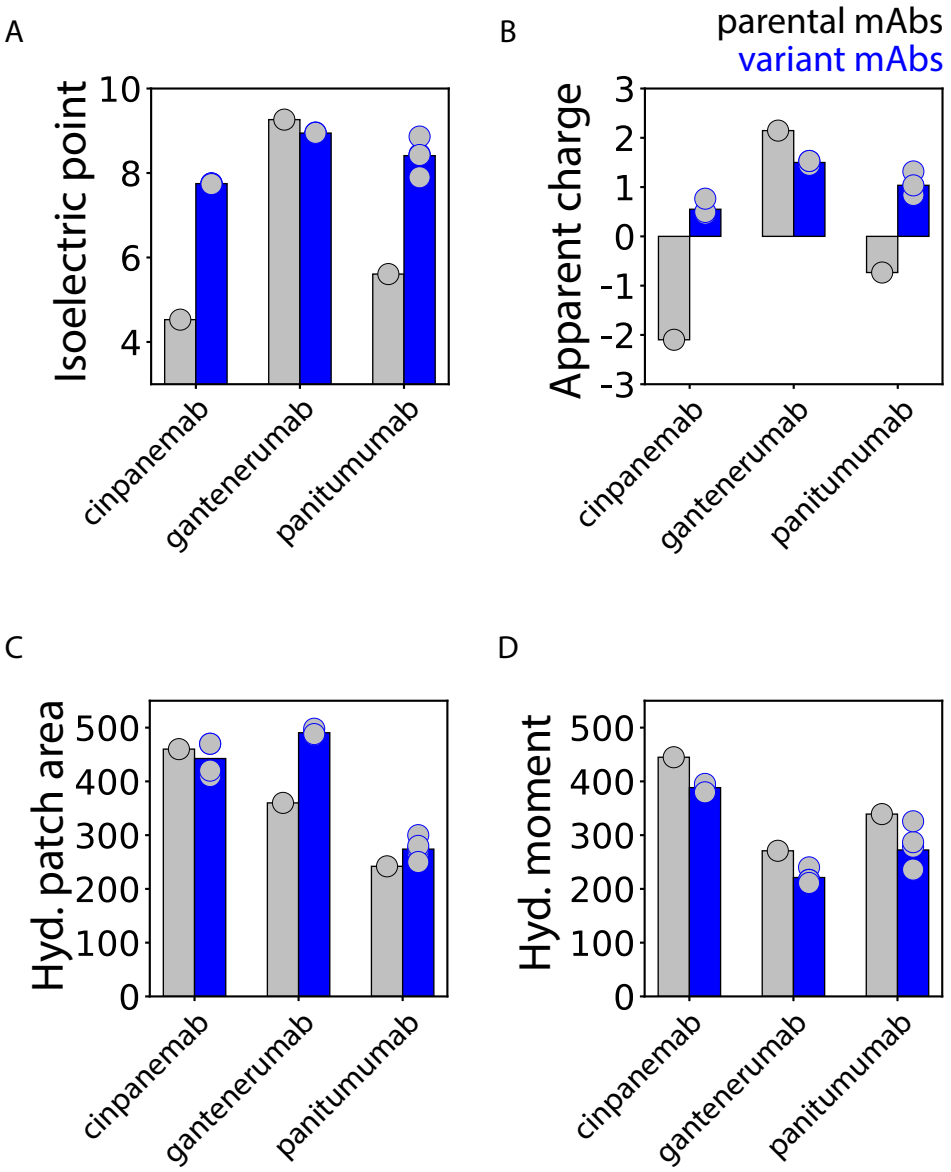
Supplementary Fig. 5 | Immunogenicity vs. Tolerance Quadrant Analysis. This graph illustrates, in two dimen-sions, the immunogenic potential of each of the original monoclonal antibodies and their modified derivatives, in terms of overall T cell epitope content (EpiMatrix score) and human-like T cell epitope content (JanusMatrix score). EpiMatrix scores (T cell epitope content) are shown on the Y axis. On the EpiMatrix scale, a score of zero represents the expectation for a random protein sequence. Anything above zero represents more putative T cell epitope content and therefore higher potential for immu-nogenicity. Scores below zero represent less putative T cell epitope content and therefore lower potential for immunogenicity. Proteins with scores above +20 are typically considered to be likely immunogenic. Immunogenic monoclonal antibodies generally run slightly lower on the scale due to the presence of regulatory T cell epitopes (Tregitopes) in their sequences. Most immunogenic anti-bodies score zero or above on this scale. Most non-immunogenic antibodies score below -20, and below -40 is preferred. The EpiMatrix analysis shows small increases in the predicted immunogenicity for the cinpanemab mutants relative to their parental antibody. Mutants of gantenerumab and pani-tumumab show only minor changes to their EpiMatrix scores.



Supplementary Fig. 6 | Measurements of the folding stability of the designed antibody mutants in the study. The antibody mutants displayed similar melting temperatures as the wild-type antibod-ies. The antibodies are defined as explained in Figure 5. Measurements are averages of two biological replicates.



Supplementary Fig. 7 | Measurements of antibody non-specific binding to ovalbumin. The antibody mutants, except for gantenerumab, displayed low non-specific binding to ovalbumin. The blue dashed line represents a published cutoff for low non-specific binding (Makowski et al., mAbs, 2021). Measurements are averages of three biological replicates.



Supplementary Fig. 8 | Physicochemical properties of antibody mutants with reduced self-association and non-specific binding. Comparison of parental antibody features with those of the designed mutants, including for (A) isoelectric point, (B) apparent charge (pH 7.4), (C) hydrophobic patch feature and (D) hydrophobic moment.