Computational and Structural Biotechnology Journal 20 (2022) 2169–2180

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Computational epitope binning reveals functional equivalence of sequence-divergent paratopes

Jarjapu Mahita [a,1,2], Dong-Gun Kim [b,1], Sumin Son [b,3], Yoonjoo Choi [c,*], Hak-Sung Kim [b,*], Chris Bailey-Kellogg [a,*]

[a] Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA
[b] Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[c] Combinatorial Tumor Immunotherapy MRC, Chonnam National University Medical School, Hwasun-gun, Jeollanam-do 58128, Republic of Korea

ARTICLE INFO

Article history:
Received 9 January 2022
Received in revised form 27 April 2022
Accepted 27 April 2022
Available online 30 April 2022

Keywords:
Epitope
Epitope binning
Paratope equivalence
Docking
Repebody
Protein binder

ABSTRACT

The therapeutic efficacy of a protein binder largely depends on two factors: its binding site and its binding affinity. Advances in *in vitro* library display screening and next-generation sequencing have enabled accelerated development of strong binders, yet identifying their binding sites still remains a major challenge. The differentiation, or "binning", of binders into different groups that recognize distinct binding sites on their target is a promising approach that facilitates high-throughput screening of binders that may show different biological activity. Here we study the extent to which the information contained in the amino acid sequences comprising a set of target-specific binders can be leveraged to bin them, inferring functional equivalence of their binding regions, or paratopes, based directly on comparison of the sequences, their modeled structures, or their modeled interactions. Using a leucine-rich repeat binding scaffold known as a "repebody" as the source of diversity in recognition against interleukin-6 (IL-6), we show that the "Epibin" approach introduced here effectively utilized structural modelling and docking to extract specificity information encoded in the repebody amino acid sequences and thereby successfully recapitulate IL-6 binding competition observed in immunoassays. Furthermore, our computational binning provided a basis for designing *in vitro* mutagenesis experiments to pinpoint specificity-determining residues. Finally, we demonstrate that the Epibin approach can extend to antibodies, retrospectively comparing its predictions to results from antigen-specific antibody competition studies. The study thus demonstrates the utility of modeling structure and binding from the amino acid sequences of different binders against the same target, and paves the way for larger-scale binning and analysis of entire repertoires.

1. Introduction

Protein binders (e.g., antibodies, nanobodies, repebodies, affibodies, DARPins, galectins, and monobodies [1–3]) are capable of specifically recognizing different target proteins with high binding affinity, making them attractive candidates for therapeutic purposes. In recent years, considerable advances have been made in the techniques and tools involved in selection and development of protein binders. In particular, adaptive immune receptor repertoire sequencing leverages next-generation sequencing technologies to characterize the sequences comprising a repertoire of B-cell receptors or antibodies [4–12], recently even including paired antibody heavy and light chains [13]. Despite such advances in rapidly discovering high-affinity binders, the process of character-

Abbreviations: Pro, Proline; RMSD, Root-mean squared deviation; AU-PRC, Area under the precision-recall curve; PCC, Pearson correlation coefficient; IL-6, Interleukin - 6; LRR, leucine-rich repeat; SARS-CoV-2, severe acute respiratory syndrome coronavirus – 2.

* Corresponding authors at: 322 Seoyang-ro, Hwasun-eup, Hwasun-gun, Jeollanam-do 58128, Republic of Korea (Y. Choi). 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea (H.-S. Kim). Dartmouth Computer Science Department, 6211 Sudikoff Laboratory, Hanover, NH 03755 USA (C. Bailey-Kellogg).

E-mail addresses: yoonjoo.choi@jnu.ac.kr, kalicuta@gmail.com (Y. Choi), hskim76@kaist.ac.kr (H.-S. Kim), cbk@cs.dartmouth.edu (C. Bailey-Kellogg).

[1] These authors contributed equally.

[2] Present address: Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA.

[3] Present address: Synthetic Biology and Bioengineering Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea.

https://doi.org/10.1016/j.csbj.2022.04.036

izing their biological functions still remains low throughput. The binding specificity of a protein binder is governed by its binding site on the target protein, also known as the 'epitope' of an antibody, a word we adopt here generically to include sites recognized by other classes of protein binder. Thus epitope mapping, the process of identifying the epitopes of high-affinity binders, is considered an essential component of understanding a binder's mechanism and function [14].

While experimental structural determination remains the gold standard for identifying epitopes and permitting deeper insights into the factors governing specific recognition of an antigen, their labor-intensive nature makes them infeasible for scaling up to large sets of interactions, as immune repertoires may yield [15]. Other methods for epitope mapping that do not involve structural determination include site-directed mutagenesis or alanine scanning mutagenesis combined with binding assays [14,16]. An alternative to epitope mapping is epitope binning, which uses competitive binding assays to sort antibodies into bins [17–21]. Unlike epitope mapping, epitope binning does not provide information about the location of the epitope on the antigen, but rather just that two binders compete with each other and thus are likely to target the same epitope (though perhaps compete due to steric hindrance distal to the epitope, conformational change by the antigen upon binding of one, etc. [22]). Subsequent characterization of a representative from each bin assists in elucidating their functions and epitopes of the whole bin. Nevertheless, experimental epitope binning, albeit higher-throughput than epitope mapping, is still limited by the size of the repertoire.

Since the sequences of protein binders encode their structures and consequently the determinants of their binding site specificity, in theory computational methods should be able to decode this information and predict where and how a set of binders with different sequences will bind an antigen. A number of different computational methods have been developed for computationally predicting binding regions (epitopes on antigens and paratopes on antibodies), applying machine learning methods based on sequence alone (e.g., [23–25]) or based on antibody structures or homology models (e.g., [26–31]). Unfortunately, the utility of purely computational prediction of epitopes and paratopes remains limited in general due to the relatively low accuracy of their predictions. To overcome the limitations of methods that are purely computational or purely experimental, and better identify epitopes, researchers have pursued methods that integrate computational predictions with experimental data. Recently, a novel approach called "dock binning" successfully localized epitopes for a panel of antibodies against glycoprotein D of herpes simplex virus, using docking models to identify putative antigenic "hot spots" on the antigen across the antibody panel, and then leveraging experimental data regarding these hot spots in order to constrain docking models and reveal each antibody's specific epitope regions [32]. This study provided a limited demonstration of the ability of dock binning to discriminate binders with different epitopes, considering only one representative antibody from each distinct "community" of epitope binders. Thus, there is still a need for a purely computational epitope binning method that not only discriminates binders having different epitopes but also clusters together sequentially-divergent binders sharing the same epitope.

In this work, we investigate the extent to which competition for binding to an epitope can be predicted directly from the amino acid sequences of a set of different but related binders, based on the sequences alone or using the sequences as the basis for models of structure and docking models which represent predicted interactions between the binders and the target. In particular, we seek to computationally analyze the functional equivalence of the binding regions, or loosely "paratopes" (as with "epitope", again adopting for general use with any protein binder the corresponding word used with antibodies), of a set of distinct binders of a common target. We take as a case study a set of interleukin-6 (IL-6) binding repebodies generated through phage display of a diversified repebody library [33]. A repebody is a binding scaffold comprised of leucine-rich repeat (LRR) modules whose β-strands mediate antigen binding [3,34]; it serves as an excellent system for the study at hand due to the ease of generating antigen-specific libraries with different binding modes driven by diversity in a relatively small set of residue positions in well-defined binding modules [35,36]. Binding competition assay experiments on IL-6 binding repebodies indicate that there are largely two different sets (bins) of binders, with one bin composed of sequence dissimilar (yet competing) binders. Computationally predicted competition revealed that although sequence similarity was in general an indicative metric for binding competition, an alternative approach was required for the set of sequence-dissimilar binders that were grouped into the same bin. Consequently, we developed Epibin, inferring functional equivalence of paratopes by leveraging the docking hypothesis that key structural features of binders mediate binding and molecular docking largely captures binding similarity. Epibin successfully classified binned variants sharing low paratope sequence identity and identified specificity-determining residues that confer functional equivalence.

2. Results

2.1. Selection and modeling of anti-IL-6 repebodies

In previous work [33], a library of repebody clones was generated by diversification of six positions in two of the leucine-rich repeat modules (hereafter called hypervariable sites 1, 2, and 3 of modules 3 and 4) and screened by phage display for IL-6 binding. Seven clones (Fig. 1) were selected for further analysis on the basis of expression, mutation variability, and binding. While the structure of one of these repebodies in complex with IL-6 (clone "D3") has been solved (PDB ID: **4J4L** [33]), the details of the interactions of the other six repebodies with IL-6 remain unknown. The sequences show diversity throughout the two variable modules (Fig. 1b). From previous structural studies [33,37], it is notable that the presence or absence of Pro in hypervariable sites (e.g. 3 of module 3, Fig. S1) induced a significant conformational change. Thus, structural models of the six repebodies were constructed accordingly, based either on the above-mentioned crystal structure for those with Pro or on a different repebody (PDB ID: **3RFS** [3]) for those without Pro (Fig. S1). Structural superposition of these two classes of repebodies consequently highlights distinct conformational differences in module 3 (Fig. 1c). ClusPro-based docking [38] of these models to the crystal structure of unbound IL-6 (PDB ID: **1ALU** [39]) yielded approximately 30 models per repebody (Fig. 1d), which were further subjected to computational binning analysis ("Epibin", Fig. 1e) as we introduce below.

2.2. Computational prediction and experimental assessment of epitope binning

Competition between repebodies was predicted based on sequence similarity (quantified as BLOSUM62 substitution scores summed across the six hypervariable positions of the repebody, Fig. 1b), binder structure-based dissimilarity (root-mean-squared deviation of the C_β atoms over the whole structure of the undocked repebody homology model, Fig. 1c), and interaction similarity (the Epibin score we introduce here, Fig. 1e) based on docking models. It was also experimentally assessed with competition binding assays. The Epibin score predicts competition for epitope binding based on pairwise similarities of complex models produced by

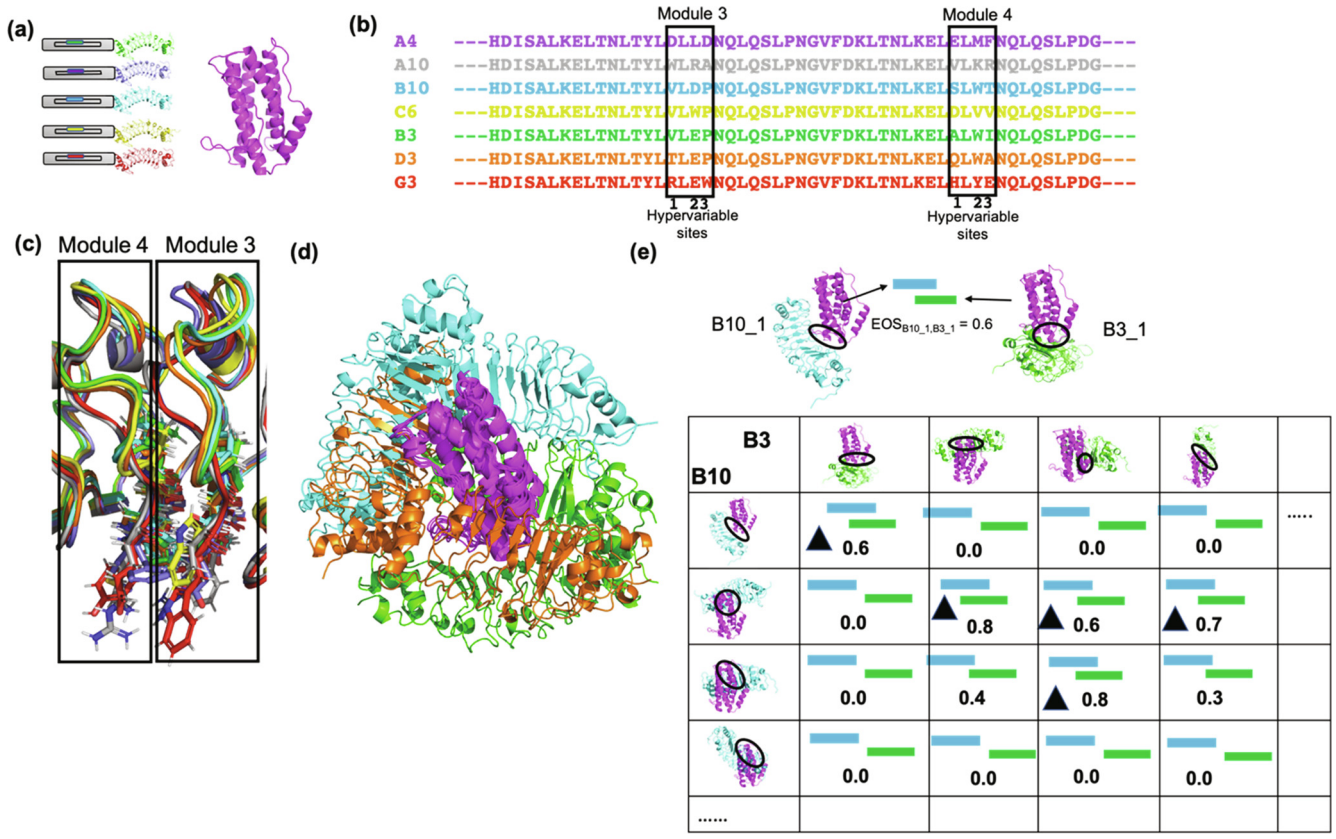


Fig. 1. Computational sequence, structure, binding, and competition analysis of IL6-binding repebody clones. (a, b) Seven repebodies were selected for analysis based on experimental properties and sequence diversity in the hypervariable sites in Modules 3 and 4. (c) The amino acid differences in the repebodies manifested as structural differences in homology models. (d) The structural differences in the repebody models led to predicted complex structure differences according to docking models (for clarity, only a few models out of ~30 per repebody are shown). (e) The docking models were analyzed by Epibin in order to predict competition for binding between pairs of repebodies. First, Epibin computes the "epitope overlap score (EOS)", the extent of common epitope residues (0: none to 1: all). In the example, B10_1 and B3_1 represent docking models of B10-IL-6 and B3-IL-6, respectively. Comparing the epitopes of these two docking models indicates that their epitopes overlap to some extent, quantified by an EOS of 0.6. Similarly, B10_2 (third row) and B3_4 (fifth column) docking models share common epitopes, yielding an EOS of 0.7. On the other hand, docking models B10_2 and B3_1 do not have any common epitope residues and hence the EOS between them is 0. Similarly, B10_4 does not share any common epitope residues with any of the B3 docking models, yielding EOS = 0.0 for the whole fourth row. Based on the set of EOSs, the overall "Epibin score" capturing competition between two repebodies is computed as the average fraction of each repebody's docking models that have a sufficiently strong competitor for the same epitope among the other repebody's docking models (i.e., EOS exceeds a threshold). In the example, the cells containing the 'Δ' symbol indicate those pairs of models for which the EOS is above the threshold of 0.5 and hence included in the fraction. Among the illustrated pairs of docking models, we see that three of the four B10 models (rows) have a strong competitor among the B3 models (columns), and all four of the B3 models have a strong competitor among the B10 models. For consistency with experimental data, the average fraction of sufficiently overlapped models is subtracted from 1 to yield the Epibin competition score, here giving a score of 1 – 0.5*(3/4 for B10 + 4/4 for B3) = 0.125.

docking. While many of the docking models are likely to be incorrect (with perhaps one near-native, some relatively close, and more well off-target [40–42]), intuitively they rely on and thus reveal complementarity between a putative paratope and putative epitope. Therefore, if the docking models of two protein binders involve many of the same regions on the target, then the binders are likely to have similar binding preferences and compete for the (unknown) true epitope. To leverage this intuition in predicting the extent of competition between binders, we devised a scoring function (Epibin score) separately considering each pair of docking models, one from each binder, and evaluating the fraction of docking models from one binder that have a sufficiently close competitor from the other, and vice versa (Eqs. 1 and 2). The fraction of models with competitors is subtracted from 1 to be comparable to experimental measurements, with Epibin score of 0 for complete competition to 1 for no competition.

Fig. 2 summarizes both the predicted competition, as heatmaps (Fig. 2a, Table S1), and the resulting binning, as dendrograms (Fig. 2b). Sequence-based analysis predicts strong competition among repebodies B10, D3, and B3 (herby referred to as the "D3 cluster") due to highly similar residues within modules 3 and 4 of these repebodies. Binning based on modelled binder structure-based dissimilarity, as estimated by RMSD, further tightly clusters A4, A10, and G3, due to the presence of Pro in the hypervariable site 2 of module 3. In contrast, Epibin predicts A4 to strongly compete with B10 and C6, to moderately compete with B3 and D3, and to weakly compete with G3. In addition, Epibin also predicts weak competition of G3 with B10 and C6. The competitive immunoassays (here symmetrized by averaging, in order to support comparison with the computational methods) showed that, consistent with Epibin but counter to sequence- and structure-based predictions, A4 indeed competes with B10 and to a lesser degree, with B3 and D3 (Fig. 2a, Fig S2). On the other hand, immunoassays revealed little competition between A4 and C6. As expected from all analyses, the repebodies within the D3 cluster also experimentally manifested high levels of competition. Finally, the experiments revealed competition between A10 and G3 that was only predicted by binder structure model-based dissimilarity analysis. In inferring clustering and bins from the patterns of competition (Fig. 2b), the experiments and all methods grouped the D3 cluster together, though Epibin brings A4 into the core of the group whereas the experiments place it just outside, and both sequence- and structure-based methods associate it most tightly with A10. Overall, the Pearson correlation coefficient (PCC) value

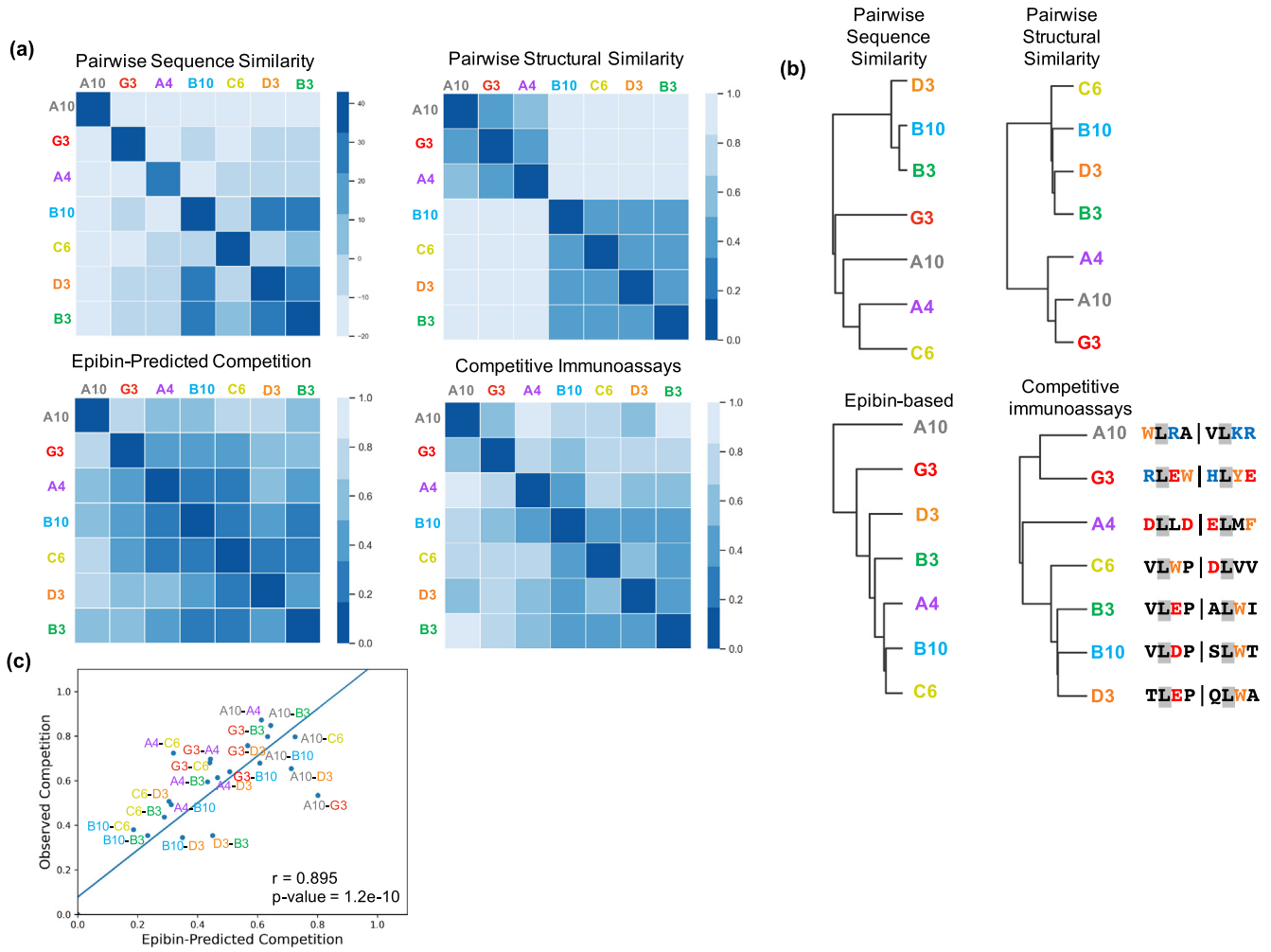


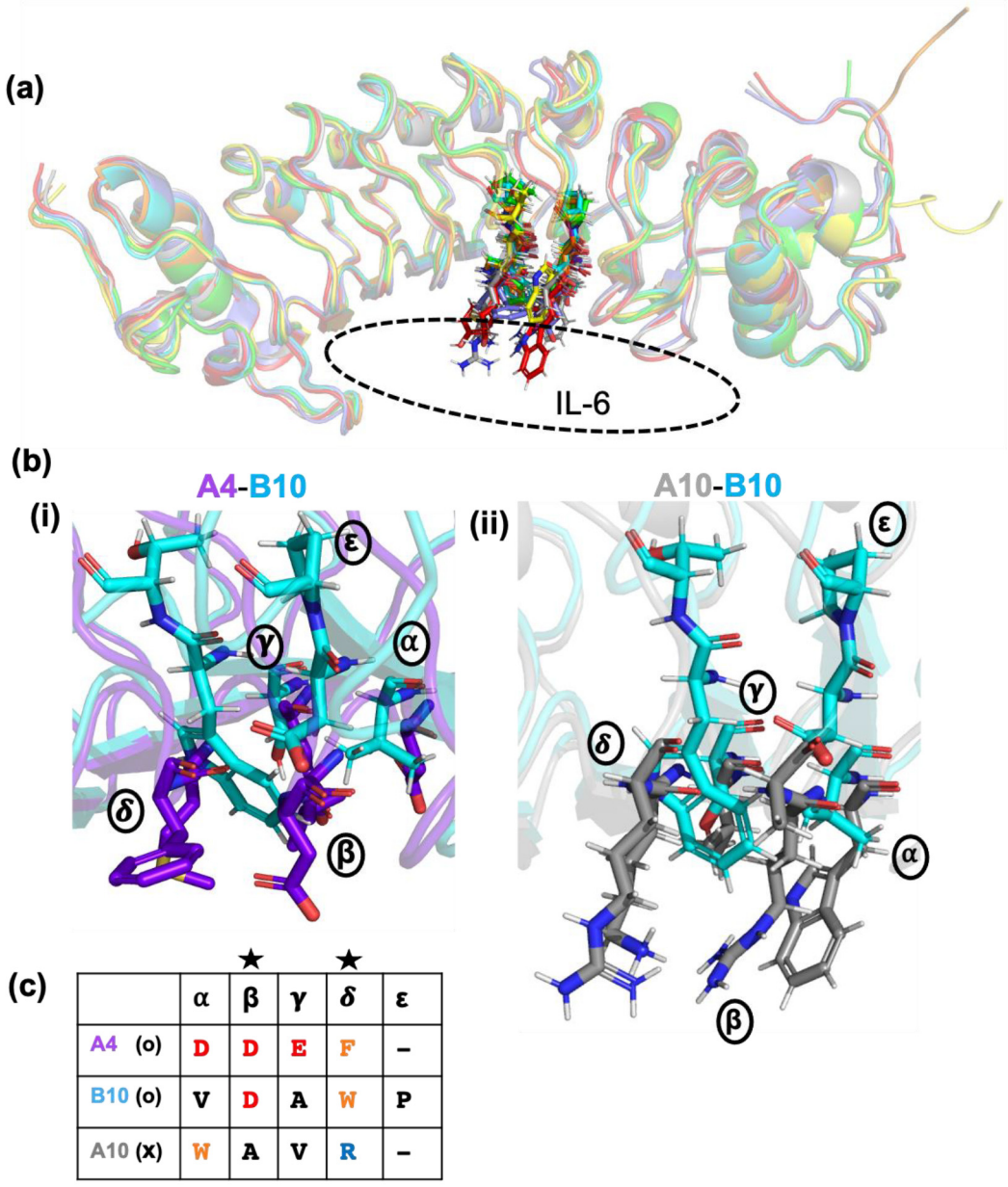
Fig. 2. Computational and experimental evaluation of repebody competition for IL-6 binding. (a) Relative competition and (b) clustering of repebodies based on pairwise sequence similarity (in terms of BLOSUM62 substitution scores across the six hypervariable positions), pairwise structural similarity as measured by C_β RMSD, pairwise complex similarity as evaluated by Epibin with an EOS threshold of 0.65, and competitive immunoassays. For both Epibin and the immunoassays, lower values indicate increased competition. (c) Correlation between Epibin-predicted and experimentally-assayed competition of the seven repebodies.

of 0.9 (Fig. 2c) indicates a strong positive correlation between Epibin-predicted competition and experimentally-observed competition, emphasizing the predictive utility of computational epitope binning.

The Epibin results presented in Fig. 2 are based on a single EOS threshold (0.65) and the complete set of ClusPro-generated docking models (a total of 30 for each repebody - IL-6 pair except for A4 - IL-6 for which ClusPro returned 24 models). To assess the impact of varying the EOS stringency and docking model completeness, we further evaluated Epibin at different EOS thresholds ranging from 0.5 to 0.75 and with sets of the top 20, top 15, top 10, and top 5 docking models each (Table S2). The PCC between Epibin-predicted competition and the observed competition when using the set of 30 models ranges from 0.69 at an EOS threshold of 0.5 to 0.94 at 0.7, indicating that this control over stringency does impact the quality of the results. For the purpose of illustrating the Epibin results in Fig. 2, we chose an EOS threshold of 0.65 (PCC of 0.9, marginally less than for the higher threshold of 0.7) since it balances accuracy over a range of test cases. Moreover, the range of PCCs over subsets of docking models varies only between 0.90 and 0.92 at this EOS of 0.65, suggesting that Epibin is robust to taking smaller subsets of the top docking models. We also evaluated the impact of the docking method, by applying Epibin to the models returned by the HADDOCK webserver [43] (a total of 400 unranked models for each repebody - IL-6 pair). However, Epibin did not fare quite as well as it did on the ClusPro docking models, with PCC ranging from 0.65 at an EOS threshold of 0.5 to a maximum of 0.78 at an EOS threshold of 0.65 (Table S2).

2.3. *"Paratope equivalent" residues drive competition between A4 and B10*

While A4 and B10 have distinct differences in sequence and modelled structures, analysis of putative interactions via Epibin was able to successfully predict their experimentally observed competition. Therefore, we sought to delineate the factors contributing to this competition. For the sake of comparison across the repebodies we label *paratope equivalent residues* that may make similar contributions to recognition of the antigen, regardless of where they are in the sequence (Fig. 3a). We base the labeling on the D3 cluster of repebodies (D3, B3, B10, and C6), which present the same paratope, and we label as 'α' the structural position contributed by their hypervariable site 1 on module 3, 'β' module 3 site 2, 'γ' module 4 site 1, and 'δ' module 4 site 2. Module 3's hypervariable site 3 is labeled 'ε', but the Pro occupying this position is pointed towards the interior of the repebody so does not participate in binding interactions for this cluster.



	α	β	γ	δ	ε
A4 (o)	D	D	E	F	–
B10 (o)	V	D	A	W	P
A10 (x)	W	A	V	R	–

Fig. 3. Competition between A4 and B10, but not A10, can be attributed to the presence of similar amino acids in "paratope-equivalent" positions. (a) Full view of the A4, B10, C6, and A10 structural models superposed onto one another, with the general location of IL-6 (from the D3 crystal structure, PDB ID: **4J4L**). (b) Zoomed-in view illustrating the positions of similar amino acids in paratope-equivalent positions of Modules 3 and 4 for (i) A4 and B10, (ii) B10 and A10. Hypervariable sites 1, 2, and 3 are represented as sticks. (c) Amino acids present in key structural positions highlight the similarity of position pairs 'β' and 'δ' (denoted by asterisks) in A4 and B10 and the presence of dissimilar residues in these positions of the non-competing repebody A10. The (o) and (x) signs indicate competition of the respective repebody with B10.

By comparing the structure of A4 to that of B10 (Fig. 3.b.i), we see that in A4 structural positions 'α' and 'γ' still derive from the same sequence positions as in B10, but in contrast 'β' is presented by module 3 site 3 instead of module 3 site 2. Similarly, 'δ' is presented by module 4 site 3 in A4 as compared to module 4 site 2 in B10. Strikingly, these structurally equivalent positions are occupied by physicochemically similar amino acids (Fig. 3c), with both repebodies using an Asp in position 'β' in A4 and B10, and the other competing repebodies B3 and D3 using Glu there (Figs. S3(a, b)). In the solved D3 - IL-6 complex structure, this Glu (Glu 128) makes electrostatic interactions with Lys 27 and Arg 30 of IL-6, and it is likely that the Asp residues are also involved in similar interactions. Likewise, the 'δ' position is filled by Phe in A4 and Trp in B10 and D3 (Fig. 3b.i; Figs. S3(a, b)), and in the solved D3 crystal structure this Trp (Trp 152) forms a hydrophobic pocket with two other aromatic residues, Phe 177 and Tyr 201. We note again that only the interaction-based prediction of competition via Epibin was able to pick up on this paratope equivalence, which uses different residue positions from A4 and B10 to fill a similar structural/functional role.

The repebodies A10 and G3 that do not compete with A4 and B10 place physicochemically different amino acids in these positions: 'β' is Ala in A10 and Trp in G3 (vs. Asp in A4 and B10), while 'δ' is Arg in A10 and Glu in G3 (vs. Phe in A4 and Trp in B10) (Fig. 3b.ii; Fig. 3c; Fig. S3.c). Thus, Epibin is able to pick up on the resulting differences in epitope preferences in predicting that they do not compete.

2.4. Experimental studies confirm the importance of paratope-equivalent residues

In order to test the hypothesis that positions 'β' and 'δ' are important for competition between A4 and the D3 cluster, we designed single-point and double-point mutant variants of A4, B10, and C6 (Table 1). Based on our hypothesis that the negatively charged amino acid in position 'β' and the aromatic residue in position 'δ' act as specificity determining positions, residues in these positions were mutated to either resemble or differ from the original physicochemical properties. In detail, the Asp residue in the 'β' position of A4 and B10 were mutated to Glu, Arg, or Lys; in the position 'δ', Phe of A4 was mutated to either Trp or Val, and the Trp of B10 was mutated to either Phe or Val. Furthermore, since B10 and C6 have different residues in the 'β' and 'δ' positions (Fig. S3.d), we reasoned that mutating the residues in these positions of C6 to mimic the interactions of B10 would increase the competition between A4 and C6 mutants (Fig. S3.e). In particular, C6 variants were constructed with the Trp residue in position 'β' mutated to Asp and/or the Val residue in position 'δ' mutated to Trp (Table 1). The competition of the variants against wild-type forms of A4, B10 and C6 were computationally predicted with Epibin as well as experimentally evaluated.

The prediction trends well (though without statistically significant correlation) with the experimental observations (Fig. 4a, Fig. S4.a). Firstly, for A4 and B10 variants, in accordance with our hypothesis, we assumed that mutating Asp in position 'β' to either Lys or Arg would greatly abrogate competition, and even binding to IL-6. As per our expectations, all of the B10 mutations to Lys or Arg in 'β' eliminated IL-6 binding. The outcomes were similar for A4 variants, where all except A4KV lost their binding to IL-6 (Fig. 4b, Fig. S4.b). In the case of A4KV, the binding affinity was maintained to a similar extent to wild-type A4 but was not able to compete with any of the wild-type binders (Fig. 4a). Mutations of Phe or Trp in 'δ' to Val removes the aromatic functional group, but since it is not a replacement by an opposing functional group (as in Asp to Lys/Arg), we expected that these mutations would likely retain the ability of these variants to bind IL-6, though with weakened strength and competition. While A4DV and A4EV maintained weak binding, complete loss of IL-6 binding was observed in B10DV and B10EV. This may originate from the fact that the wild-type sequence in 'δ' for A4 is Phe and B10 is Trp. In this sense, we assume that B10 undergoes greater reduction in the size of the functional group in 'δ' when mutated to Val and therefore manifests a larger decrement in binding compared to A4. In line with our hypothesis, experimental evaluation showed that A4EV competed to a lesser extent with wild-type binders for IL-6 binding. Other than A4EW, the physicochemically-similar mutations (i.e., A4DW, B10DF, and B10EF), exhibited similar trends of predicted and observed competition.

Next, we set out to investigate whether the introduction of Asp in the 'β' position of C6 and Trp in the 'δ' position would lead to increased competition with A4. All of the C6 mutants were predicted to compete with the wild-type C6 to a similar extent. However, we found that the mutation in position 'β' to Asp induced a formation of oligomeric species for both C6DV and C6DW (Fig. S4c.(iii, iv)), which is likely the main reason for these two variants displaying significantly weaker binding to IL-6. Nevertheless, increased competition for C6WW against both A4 and B10 wild-types indicates that an aromatic functional group in 'δ' is indeed a key specificity-determining residue (Fig. 4a). Overall, these results support Epibin's ability to uncover structural/functional equivalence in the paratope, thus outperforming sequence-based or binder structure-based predictions in predicting competition.

Table 1
List of designed variants of A4, B10 and C6. Mutations made in repebody modules 3 and 4 (position 'β' and/or position 'δ') are highlighted in bold.

Module 3	Module 4	Variant
DLLD	ELMF	A4DF (wild-type)
DLLD	ELM**W**	A4W
DLL**E**	ELM**W**	A4EW
DLLD	ELM**V**	A4V
DLL**E**	ELM**V**	A4EV
DLL**K**	ELMF	A4K
DLL**K**	ELM**V**	A4KV
VLDP	SLWT	B10DW (wild-type)
VLDP	SL**F**T	B10F
VL**E**P	SL**F**T	B10EF
VLDP	SL**V**T	B10V
VL**E**P	SL**V**T	B10EV
VL**R**P	SLWT	B10R
VL**R**P	SL**V**T	B10RV
VLWP	DLVV	C6WV (wild-type)
VL**D**P	DLVV	C6D
VL**D**P	DL**W**V	C6DW
VLWP	DL**W**V	C6W

2.5. Localizing epitope regions

While investigation of competition gives an idea about which repebodies are likely to bind to the same epitope, along with their paratope-equivalent residues, it does not provide any information about the location of the epitopes. The observed competition between A4 and B10 suggests that their epitopes are common or located in close proximity to one another. In addition, based on the high sequence identity between the members of the D3 cluster of repebodies (B3, B10, C6, and D3), it is conceivable that the D3 epitope is shared by the other members of the D3 cluster. Also, while immunoassays showed competition between A10 and G3, Epibin did not predict any competition between these two repebodies. To better understand the mechanisms at work here, we sought to localize the epitope on IL-6 for each of the six repebodies. We made use of EpiScope [44], a method that views antigen–antibody docked models as hypothesized binding modes and designs sets of antigen variants to test these hypotheses. If a docking model is correct, then unfavorable mutations in the associated interface (e.g., switching amino acid charge, size, hydrophobicity, etc.) should disrupt binding. EpiScope optimizes small panels of such mutational variants to efficiently test the various docking models, using computational protein design to identify combinations of mutations likely to disrupt putative binding while maintaining antigen stability. EpiScope has been successfully applied to localize the binding site of many protein binders with diverse scaffolds [35,44–46]. Here, EpiScope was applied separately to each repebody, designing each repebody-specific IL-6 triple-mutants of which their binding disruption was examined experimentally, and thereby localizing the binding region.

Table 2 shows the list of mutated sequence for each repebody. EpiScope designed a minimal set of triple-mutants predicted to effectively disrupt binding regardless of which docking model is most accurate. In the case of the D3 cluster of repebodies (i.e., B3 and B10), which have high sequence similarity, the available IL-6-bound crystal structure was directly used in place of docking models to guide the design. The residues that are mutated in these IL-6 designs are highlighted in Fig. 5a. Overall, three distinct candidates for the epitope regions can be discerned: one containing residues Arg 24, Lys 27, Arg 30, and Tyr 31 (colored green in Fig. 5a), one Ala 38, Arg 40, Lys 41, Asn 45, Met 49, and Ser 52 (colored brown in Fig. 5a), and one Ala 68, Phe 74, and Glu 80 (colored blue in Fig. 5a). The first epitope region overlaps with the crystal structure-based design that abrogates D3 binding; thus, the bind-

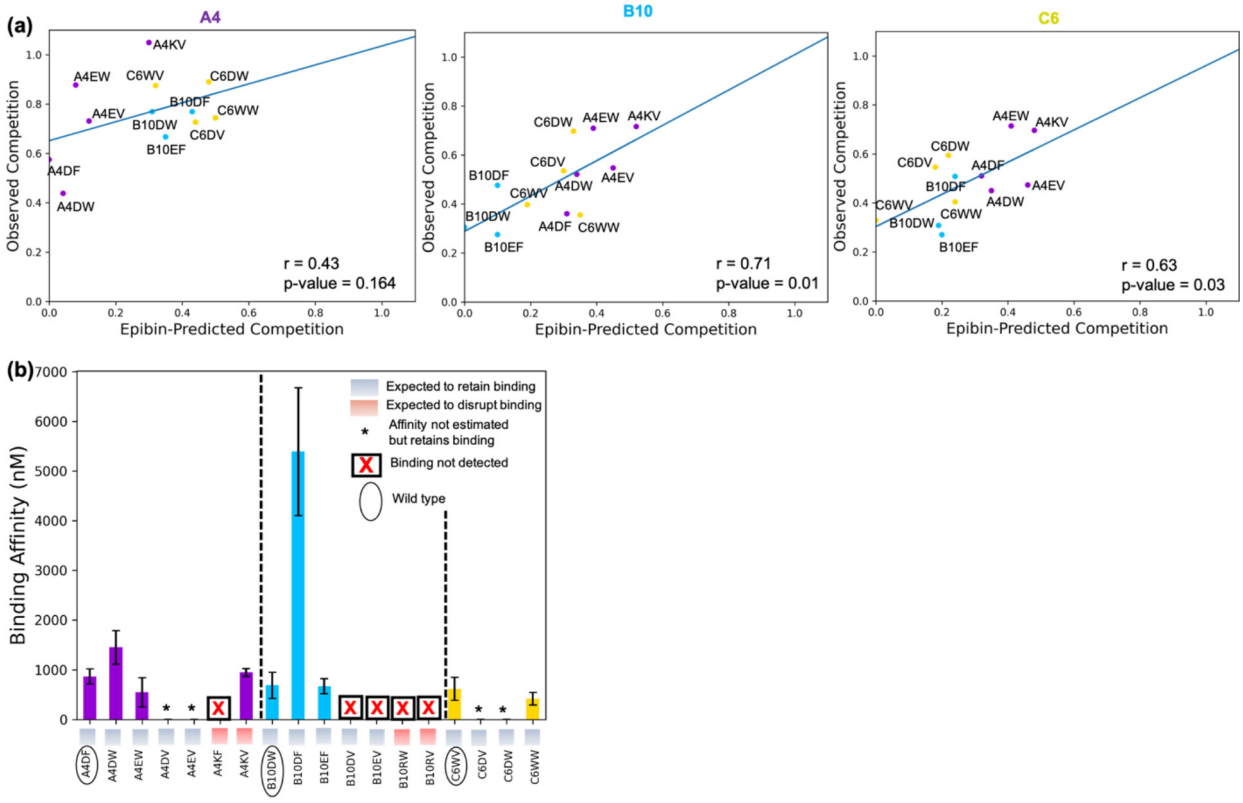


Fig. 4. Competitive immunoassays and isothermal titration calorimetry (ITC) results confirm the significance of paratope-equivalent positions 'β' and 'δ'. (a) Correlation between Epibin-predicted competition (x-axis) and corresponding observed competition (y-axis) of all designed variants competing against the wild-type forms A4, B10, and C6 (separate panels by wild-type). Pearson's correlation coefficient is denoted as r. (b) Binding affinity measurement of the wild-type and variant forms of A4, B10, and C6 against IL-6 as measured by ITC. Bars show the means and SDs over three replicates.

Table 2
EpiScope-generated triple-mutant variants of anti-IL-6 repebodies to localize epitope regions on IL-6.

Repebody	Triple-mutant variants
A10	A50T|F56L|E62K, K9E|Y13S|A20E, N27D| M31K|S34N
A4	Y13S|R22S|K23E, R6K|K9E|Y13S, R22S|K23D|N27E
B10	R6Q|R12K|Y13S, N27D|M31K|S34D, Y13S|R22K|K23E
B3	R6Q|R12K|Y13S, N27D|N30D|M31K, R12K|Y13S|K23E
C6	N27E|M31V|S34D, R6Q|R12K|Y13S, R12K|R22K|K23E
G3	R22K|N27E|M31K, R6K|K9E|Y13S, R12K|Y13S|R22S
4J4L (D3 Clone)	R6K|K9E|R12K, R6Q|K9E|R12K, L1G|R6K|K9E, L1G|R6Q|R12K, L1G|R6L|R12K

ing of the D3 cluster and A4 to designs spanning this region were expected to be disrupted. The third region is unique for A10, which implies that A10 produces docking models different from those produced by the other repebodies.

Binding disruption was evaluated via ELISA testing repebody binding against the respective designed IL-6 triple mutants vs. that against wild-type IL-6 (Fig. 5b, Fig. S5). As expected, B3, B10, and C6 were all revealed to interact with IL-6 in a binding mode similar to that of D3, as inferred by binding disruption of the triple-mutant in the D3 binding interface. Surprisingly, A4 was shown to interact predominantly with the second epitope region, whose mutations largely disrupted its binding, while mutations in the first epitope region resulted in intermediate binding disruption. This implies that although A4 shares similar functional paratope equivalence with the D3 cluster of repebodies (position **'β'** and **'δ'** in Fig. 3), they do not employ identical binding modes as their backbone alignments may differ. This could also be one of the reasons A4 was more resistant than B10 to the mutation of the 'δ' position Phe to Val. For A10, binding was disrupted only for the third epitope region, located opposite to the D3 epitope, and for G3 none of the variants resulted in any reduction in binding. Since in the competitive ELISA, A10 and G3 were shown to compete, we also tested binding of G3 to the IL-6 triple-mutant variants designed for A10. G3 was found to retain binding to these variants, which suggests that the G3 epitope is likely located in the vicinity of the A10 epitope and the observed competition between them could be due to steric clashes while attempting to bind their respective epitopes. This further illustrates that experimental competition assays cannot determine whether the competition is due to an overlap of the epitope or, as in this case, steric hindrance. Therefore, for the purpose of selecting and developing therapeutics, it may be very beneficial to leverage the information provided by Epibin regarding whether the binders actually share an epitope or not.

2.6. General applicability of Epibin: antibody binning

The ability to group entire repertoires of antibody sequences by predicted epitope specificity could provide large-scale functional insights into the effects of infection and vaccination, as well as the results of antigen-specific library screening. While our main focus here is on repebodies against IL-6 due to the availability of high-quality experimental data and the ease of conducting follow-up experiments, we were also interested in assessing how well Epibin is able to bin groups of antigen-specific antibodies. In order to gain preliminary insights, we investigated three different sets of antibodies for which experimental cross-blocking data is available: twelve anti-D8 antibodies [47], ten anti-Pfs25 antibodies [48], and nine anti-SARS-CoV-2 antibodies [49]. Following an analogous approach to that implemented for repebodies, competition

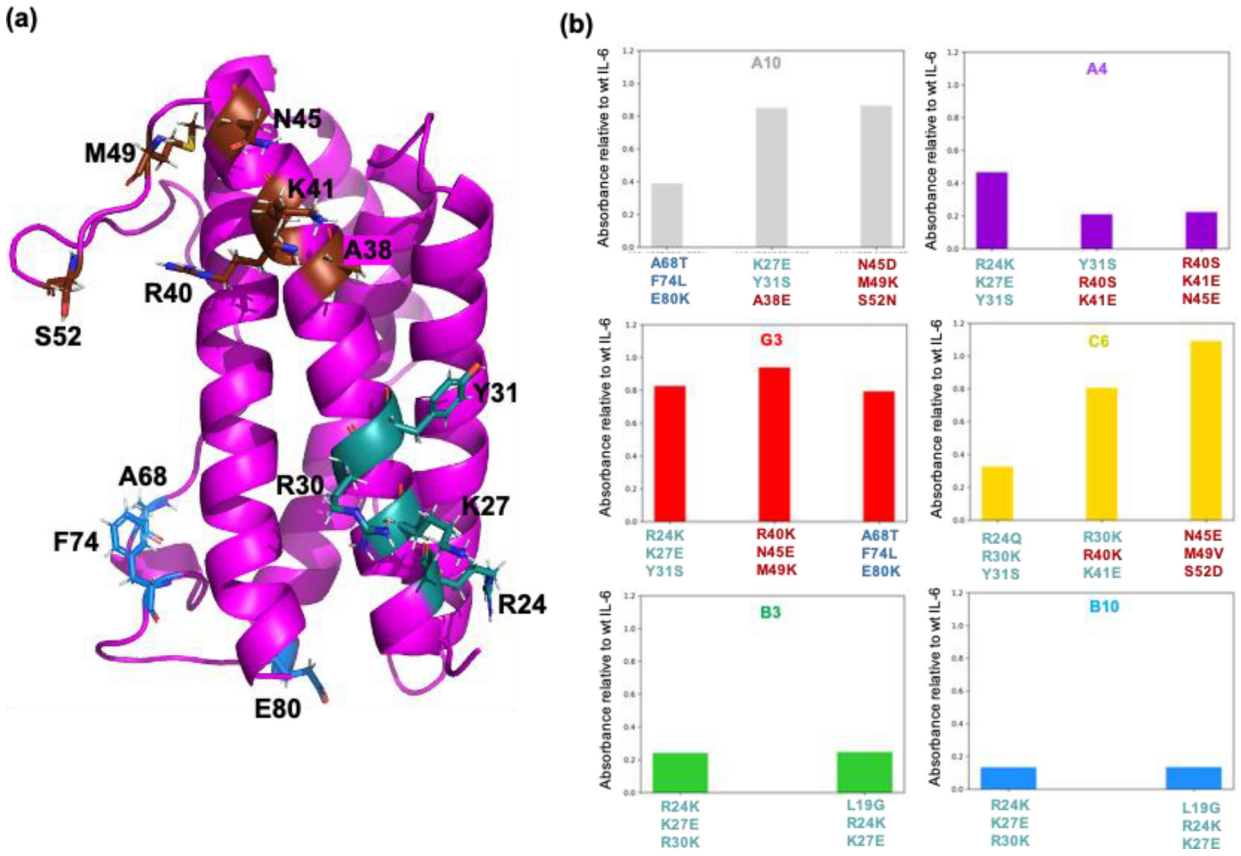


Fig. 5. Mutational studies confirm that epitopes of A4 and B10 partially overlap. Triple-mutant variants of IL-6 were designed by Episcope [44] to localize repebody epitopes by disrupting hypothesized binding interfaces while maintaining IL-6 stability. (a) Residues selected by Episcope for mutation mapped onto the structure of IL-6 (PDB ID: **1ALU**) and colored by general regions. (b) Bar plots representing the normalized binding of IL-6 variants, relative to wild type IL-6, to each of the six repebodies. Colors of mutated residues refer to colors of regions in panel (a). Values are the means of three replicates.

between the antibodies was predicted based on sequence identity among the CDR loops, which mainly drive antibody recognition of the antigen, structural similarity of the variable regions, and Epibin-based comparison of docking models. Furthermore, for Epibin evaluation, both ClusPro and HADDOCK models were tested, using either the full ClusPro set of ~25–30 models or just the top 5, 10, 15, or 20, and the EOS threshold was varied from 0.45 to 0.7. To evaluate the ability of the different computational binning methods to correctly classify pairs of antibodies as either competing or non-competing, we calculated the area under the precision-recall curve (AU-PRC) evaluating agreement between predicted and observed competition (as classified in the original publications) over the range of prediction probability thresholds. We note that AU-PRC is preferred to area under receiver operating characteristic curve, AU-ROC, for imbalanced datasets like these.

We observed that overall, the best results were obtained with the top 20 ClusPro models at an EOS of 0.45 (Table S3). In the case of the anti-Pfs25 antibodies, CDR sequence identity yielded 0.80 AU-PRC and Epibin 0.78, while structural similarity of the antibody variable region only obtained 0.56 (Table S3). For the anti-SARS-CoV-2 dataset, Epibin substantially outperformed the others, with 0.60 compared to 0.32 for sequence and 0.25 for antibody variable region structure (Table S3). However, for the anti-D8 dataset, structural similarity of the antibodies as gauged from RMSD measurements was by far the best score, yielding 0.91 AU-PRC, compared to 0.40 for sequence similarity and 0.60 for Epibin (Table S3). Epibin results using HADDOCK models were not quite as good as those using ClusPro models (Table S3; note that the server failed to return results for the anti-D8 Abs).

Although Epibin demonstrated promising performance on antibodies, the modest results when compared to its performance on repebodies could be attributed to other factors such as the inherent nature of the datasets where the antibodies themselves could possibly have widely different epitope specificities with little room for competition, or have been chosen based on a sequence identity cutoff. Thus, despite having scope for improvement, these initial results also demonstrate the general utility of Epibin for binning large sets of antibodies based on available information about the antibody sequences and antigen.

3. Discussion

Epitope binning facilitates the selection of diverse antigen-binding candidates from a large pool by grouping into bins those with shared features. While experimental binning methods have scaled to assess competition between numerous clones, next-generation sequencing methods go much further, identifying thousands or even millions of different clones in repertoires [12,50]. This gap begs for computational methods that can characterize the different sets of binding functions within a repertoire, based on sequence information alone, without needing to isolate and experimentally evaluate the different clones. The study presented here evaluates the extent to which comparison of sequences, modeled structures, or modeled interactions, leads to binning of a set of seven anti-IL-6 repebodies in a manner consistent with experimentally-observed competition. We show that the information provided by our Epibin computational epitope binning method was beneficial not only in distinguishing important com-

petition, but also in pinpointing the paratope-equivalent residues contributing to the binding specificity of each bin.

For both repebodies and antibodies, the accuracy of a computational predictor naturally depends on the quality of the models. Here, binning predictions are based on docking models which in turn are influenced by the nature of the interacting proteins. For example, protein interactions involving flexible binding sites or significant conformational changes are particularly challenging. The docking models further depend on the initial models of the individual proteins. In the case of repebodies, the presence of a Pro in one of the binding loops may affect the quality of the model, while in the case of antibodies, CDR-H3 may be difficult to model correctly. Furthermore, the antigen structure used in docking may not capture the same state as that probed experimentally, e.g., if there are post-translational modifications such as glycosylation which is common in viral antigens. Despite these limitations, with the methods for modeling proteins and interactions continuing to improve and the databases upon which they build continuing to expand, computational epitope binning promises to enable large-scale investigation of functional diversity across repertoires of antigen-specific binders.

4. Materials and methods

4.1. Generation and selection of repebodies

The generation and selection of a panel of diverse anti-IL-6 repebodies was previously described [33]. All 42 clones were compared against each other and those with more than 4 common sequences in the hypervariable region were discarded from the set. Among the 11 clones with sequence variance, only those with sufficient and comparable binding intensity were chosen through ELISA (data not shown).

4.2. Structural modelling and docking of repebodies against IL-6

Due to the structural importance of Pro, repebody homology models were based on the D3 crystal structure (PDB ID: **4J4L**, [33]) for those containing Pro in the hypervariable site 3 of module 3 (B3, B10, and C6) and on a different repebody crystal structure (PDB ID: **3RFS** [3]) for those without Pro (A4, A10, and G3). Twenty models per repebody were generated using MODELLER v9.12 [51] and the best model was selected based on the DOPE score [52]. Models were further energy-minimized by the Tinker (ver. 6) molecular dynamics software [53] using the GB/SA implicit solvent model [54] with the AMBER99sb force field parameter [55]. The structure of unbound IL-6 (PDB ID: **1ALU**) [39] contains a few missing residues which were filled using MODELLER and energy-minimized as described above. The modeled repebodies were docked on the unbound structure of IL-6 using two docking webservers – ClusPro [38] and HADDOCK [43]. When utilizing the ClusPro webserver, the Antibody Mode scoring function [56] was employed, using residues making up the concave portion of the repebody to guide the docking. The ClusPro webserver generated 24–30 models for each repebody – IL-6 pair. For the HADDOCK webserver, residues comprising the concave portion of the repebody and all surface residues of IL-6 were used to guide the docking. The HADDOCK webserver generated 400 unranked models for each repebody – IL-6 pair.

4.3. Computational prediction of competition

Repebodies were evaluated for competition based on comparisons of their sequences, their modeled structures, and their modeled interactions with the antigen.

4.3.1. Sequence-based similarity

Repebody sequences vary only in six positions that are restricted to Modules 3 and 4. All-against-all pairwise BLOSUM62 substitution scores, used as a measure of similarity, was calculated on the basis of these six hypervariable positions.

4.3.2. Binder structure-based dissimilarity

All-against-all pairwise root mean square deviations (RMSD) of C_β atoms were calculated for the modelled structures of the six repebodies along with the D3 crystal structure (Fig. 1c). For the antibodies, C_β-RMSD was calculated on the variable regions of heavy and light chains.

4.3.3. Complex model-based competition score: Epibin

The Epibin score for a pair of protein binders integrates information about their docking models against a common target to assess how likely it is that the binder will compete for the same epitope on that target. We define an antigen residue to be in the epitope region according to a particular docking model if any of its non-hydrogen heavy atoms are within 6 Å of any of the binder's heavy atoms in that docking model. We compute the "epitope overlap score" (EOS) (Fig. 1e) for a pair of docking models based on the similarity of their two sets of "epitope" residues defined in the models. Let A and B be two docking models and $e(\cdot)$ be a function that gives the epitope residues in a docking model as described above. Then the EOS is defined as the Jaccard index:

$$EOS(A,B) = \frac{|e(A) \cap e(B)|}{|e(A) \cup e(B)|} \quad (1)$$

For the Epibin score, we compute the fraction of one binder's docking models that have competitors from the other binder. Let R and S be two sets of docking models and $I\{\cdot\}$ the indicator function yielding 1 if a predicate is true and 0 if it is false, based on the overlap threshold θ. Then the Epibin score is defined as follows:

$$Epibin(R,S) = 1 - 0.5\left(\frac{\sum_{A\in R} I\{\exists_{B\in S} EOS(A,B) \geq \theta\}}{|R|} + \frac{\sum_{B\in S} I\{\exists_{A\in R} EOS(B,A) \geq \theta\}}{|S|}\right) \quad (2)$$

A score of "0" indicates the two sets of docking models are most likely overlapped, whereas "1" means no docking models share a sufficient number of epitope overlaps according to the epitope overlap threshold (θ). The results presented here use $\theta = 0.65$ except when examining the effect of different values of θ on the Epibin scores for the repebodies and antibodies as indicated in the Results section.

4.4. Binning

Repebodies were hierarchically clustered (calculated by average linkage) based on one of these similarity/dissimilarity scores using the clustermap function of the Python data visualization software Seaborn [57].

4.5. Protein expression and purification

Human interleukin-6 (IL-6; amino acids 19–184) was cloned into a pET21a vector with 6xHis-tag and a thrombin cleavage sequence inserted at the N-terminus of the IL-6 gene. Repebody genes were selected from the previous report [33] and synthesized by Integrated DNA Technologies. Repebody genes were cloned into a pET21a expression vector with either a myc tag or a flag tag attached to the C-terminus. All cloned genes were transformed to Origami B(DE3) cells for expression. Cells were cultured at 37 °C, 200 rpm using LB media to OD 0.6 and induced with 0.5 mM IPTG.

Induced cells were harvested by centrifuge at 4,000 g, after 20 h of culture at 18 °C, 200 rpm. Harvested cells were lysed through sonication and centrifuged at 13,000 rpm for 1 h. The supernatant was filtered, and affinity purified with Ni-NTA (Qiagen). Proteins were further purified with 20 mM Tris-HCl (pH 8.0), 50 mM NaCl buffer through size exclusion chromatography (AKTA FPLC, GE Healthcare).

4.6. Experimental epitope binning by competitive ELISA

Competitive ELISAs were conducted with the seven repebodies. Human IL-6 (10 μg/mL) was immobilized to 96-well maxisorp plate (SPL Life Sciences) overnight at 4 °C. Plates were blocked with 2% PBST-BSA for 2 h at 4 °C. Purified repebody A (myc tag) was mixed with another repebody B (flag tag) in molar ratios of 1:0, 1:1, 1:2 and 1:4 to a final concentration of repebody A of 100 μg/mL. After washing the 96-well plate with PBST, the repebody mixture was treated and incubated for 1 h at room temperature. Plates were washed three times before addition of anti-myc tag antibody-HRP conjugate in 1:400 dilution (sc-40 HRP; Santa Cruz Biotechnology). After 1 h of incubation with antibody at room temperature, tetramethylbenzidine solution (Sigma) was added to each well, and the reaction was stopped immediately by adding 1 M sulfuric acid. The results were obtained by scanning absorbance at 450 nm using Infinite M200 plate reader (Tecan).

The experimental competition score was calculated as the ratio of the absorbance taken at 1:0 (myc:flag) to the absorbance observed at 1:4. Competition between two repebodies was considered to be present if this ratio was ≤ 0.5. In order to render the results directly comparable with the computational prediction, which are necessarily symmetric, the A:B score and B:A score were averaged to yield a symmetric competition score (Table S1).

4.7. Epitope localization of repebodies

Repebody epitopes were localized by application of the computational-experimental EpiScope method [44]: design IL-6 variants mutated so as to disrupt docking model interface regions, and experimentally evaluate the effects on binding of the mutations in order to test the hypothesized binding modes.

4.7.1. Design

The docking models described above were provided as input to EpiScope, which designed for each repebody a set of repebody-specific triple-mutant IL-6 variants predicted to maintain IL-6 stability while disrupting each of the binding interfaces present in the models. There were a total of 12 triple-mutant variants, three each for A4, A10, C6, and G3. Due to their high sequence similarity with D3, clones B3 and B10 were expected to resemble the binding mode of D3. Thus, triple-mutant IL-6 variants based on the crystal structure of D3 in complex with IL-6 (PDB code 4J4L) were designed and tested prior to EpiScope. EpiScope was not applied to B3 and B10 as the crystal structure-based triple-mutant successfully disrupted binding of the two clones.

4.7.2. Experimental evaluation of repebody binding against IL-6 triple-mutant variants

The designed variants of IL-6 with triple point mutations were cloned, expressed, and purified in the same methods as wild-type human IL-6. For the binding assays, 10 μg/mL of each IL-6 variant was immobilized onto 96-well plates by incubation overnight at 4 °C. The plates were blocked with 2% PBST-BSA solution for 2 h at 4 °C. Following washing with PBST, each repebody and corresponding IL-6 variants were treated in a range of concentrations and incubated for 1 h at room temperature. The subsequent steps from antibody treatment to absorbance scanning were identical to those described above for competitive ELISA.

4.8. Binding affinity measurement

The binding affinity was determined through ITC at 25 °C (iTC200 system; Microcal). Repebody was prepared in 0.2 mM, and IL-6 in 0.02 mM. 2 μL of repebody was injected twenty times to IL-6 in an isothermal chamber. Integrated peaks were fitted to a one-site binding model, and the dissociation constant was calculated using the Origin program.

4.9. Structural modelling, docking, and computational binning of antibodies

Anti-D8, anti-Pfs25 and anti-SARS-CoV-2 antibodies whose structures were unavailable were modelled using ABodybuilder [58] and the top-ranked models were chosen for docking to the crystal structure of the vaccinia virus D8 antigen (PDB ID: **4E9O**), Pfs25 antigen (PDB ID: **6AZZ**) and the RBD domain of SARS-CoV-2 (PDB ID: **6XDG**), respectively using either the ClusPro docking server with the Antibody Mode scoring function or the HADDOCK webserver [43]. Residues making up the CDRs were used to guide the docking in the case of ClusPro, while for HADDOCK, the antibody CDR residues and all surface-exposed residues of the respective antigen structures were used to guide the docking. The HADDOCK webserver failed to return results for the anti-D8 antibodies, so only results for the other two are provided here. Epibin was then implemented on the antibody docking models in the same manner as that done for the repebody-IL-6 docked models.

Accession numbers

Protein Data Bank accession numbers:

PDB ID: 4J4L, PDB ID: 3RFS, PDB ID: 1ALU, PDB ID: 4E9O, PDB ID: 6AZZ, PDB ID: 6XDG.

Glossary

Epitope: the region on an antigen recognized by an antibody, here generalized to include the region on any target protein recognized by any target-specific protein binder.

Epitope binning: grouping antigen variants into sets sharing common epitopes by which antibodies recognize them, here generalized to the grouping of a set of any protein target variants based on common binding regions by which target-specific protein binders recognize them.

Paratope: the region on an antibody recognizing an antigen, here generalized to include the region on any target-specific protein binder recognizing its target.

Repebody: a binding scaffold comprised of leucine-rich repeat modules whose β-strands mediate target binding.

CRediT authorship contribution statement

Jarjapu Mahita: Investigation, Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Dong-Gun Kim:** Investigation, Methodology, Validation, Formal analysis, Visualization, Writing – review & editing. **Sumin Son:** Validation. **Yoonjoo Choi:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Hak-Sung Kim:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Chris Bailey-Kellogg:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Global Research Laboratory Program (NRF-2015K1A1A2033346) to Hak-Sung Kim, the Combinatorial Tumor Immunotherapy MRC grant of the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2020R1A5A2031185) to Y.C., and NIH R01 2R01GM098977 to CBK.

The authors also thank Arij Elfaki (Dartmouth) for her earlier methodology development on which Epibin is based, as well as the Research Computing (Information Technology and Consulting) group at Dartmouth for access to and maintenance of the Discovery cluster.

Competing interests

The authors declare no competing interests.

Data statement

All docking models, intermediate files, processed experimental data, and analysis scripts developed for the studies described in this paper have been deposited in a GitHub repository which can be accessed at https://github.com/jmahita/Epibin. The experimental data for the antibody studies was obtained from the supplementary material of the respectively cited publications.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.04.036.

References

[1] Harmansa S, Affolter M. Protein binders and their applications in developmental biology. Dev 2018. https://doi.org/10.1242/dev.148874.
[2] Aguilar G, Vigano MA, Affolter M, Matsuda S. Reflections on the use of protein binders to study protein function in developmental biology. Wiley Interdiscip Rev Dev Biol 2019. https://doi.org/10.1002/wdev.356.
[3] Lee SC, Park K, Han J, Lee JJ, Kim HJ, Hong S, et al. Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. Proc Natl Acad Sci U S A 2012. https://doi.org/10.1073/pnas.1113193109.
[4] Marks C, Deane CM. How repertoire data are changing antibody science. J Biol Chem 2020. https://doi.org/10.1074/jbc.REV120.010181.
[5] Rouet R, Jackson KJL, Langley DB, Christ D. Next-generation sequencing of antibody display repertoires. Front Immunol 2018. https://doi.org/10.3389/fimmu.2018.00118.
[6] Robinson WH. Sequencing the functional antibody repertoire. Nat Rev Rheumatol 2014.
[7] Tian X, Li C, Wu Y, Ying T. Deep mining of human antibody repertoires: concepts, methodologies, and applications. Small Methods 2020. https://doi.org/10.1002/smtd.202000451.
[8] Chaudhary N, Wesemann DR. Analyzing immunoglobulin repertoires. Front Immunol 2018. https://doi.org/10.3389/fimmu.2018.00462.
[9] Friedensohn S, Khan TA, Reddy ST. Advanced methodologies in high-throughput sequencing of immune repertoires. Trends Biotechnol 2017. https://doi.org/10.1016/j.tibtech.2016.09.010.
[10] Wardemann H, Busse CE. Novel approaches to analyze immunoglobulin repertoires. Trends Immunol 2017. https://doi.org/10.1016/j.it.2017.05.003.
[11] Ravn U, Didelot G, Venet S, Ng KT, Gueneau F, Rousseau F, et al. Deep sequencing of phage display libraries to support antibody discovery. Methods 2013. https://doi.org/10.1016/j.ymeth.2013.03.001.
[12] Briney B, Inderbitzin A, Joyce C, Burton DR. Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature 2019. https://doi.org/10.1038/s41586-019-0879-y.
[13] Wang B, Dekosky BJ, Timm MR, Lee J, Normandin E, Misasi J, et al. Functional interrogation and mining of natively paired human v H: V L antibody repertoires. Nat Biotechnol 2018. https://doi.org/10.1038/nbt.4052.
[14] Nilvebrant J, Rockberg J. An introduction to epitope mapping. Methods Mol Biol 2018. https://doi.org/10.1007/978-1-4939-7841-0_1.
[15] Toride King M, Brooks CL. Epitope mapping of antibody-antigen interactions with X-ray crystallography, in. Methods Mol Biol 2018. https://doi.org/10.1007/978-1-4939-7841-0_2.
[16] Moreira GMSG, Fühner V, Hust M. Epitope mapping by phage display. Methods Mol Biol 2018. https://doi.org/10.1007/978-1-4939-7447-4_28.
[17] Brooks BD, Miles AR, Abdiche YN. High-throughput epitope binning of therapeutic monoclonal antibodies: Why you need to bin the fridge. Drug Discov Today 2014. https://doi.org/10.1016/j.drudis.2014.05.011.
[18] Sivasubramanian A, Estep P, Lynaugh H, Yu Y, Miles A, Eckman J, et al. Broad epitope coverage of a human in vitro antibody library. MAbs 2017. https://doi.org/10.1080/19420862.2016.1246096.
[19] Abdiche YN, Miles A, Eckman J, Foletti D, Van Blarcom TJ, Yeung YA, et al. High-throughput epitope binning assays on label-free array-based biosensors can yield exquisite epitope discrimination that facilitates the selection of monoclonal antibodies with functional activity. PLoS ONE 2014. https://doi.org/10.1371/journal.pone.0092451.
[20] Chan BM, Badh A, Berry KA, Grauer SA, King CT. Flow cytometry-based epitope binning using competitive binding profiles for the characterization of monoclonal antibodies against cellular and soluble protein targets. SLAS Discov 2018. https://doi.org/10.1177/2472555218774334.
[21] Jia X-C, Raya R, Zhang L, Foord O, Walker WL, Gallo ML, et al. A novel method of Multiplexed Competitive Antibody Binning for the characterization of monoclonal antibodies. J Immunol Methods 2004;288:91–8. https://doi.org/10.1016/j.jim.2004.02.010.
[22] Abdiche YN, Yeung AY, Ni I, Stone D, Miles A, Morishige W, et al. Antibodies targeting closely adjacent or minimally overlapping epitopes can displace one another. PLoS ONE 2017. https://doi.org/10.1371/journal.pone.0169535.
[23] Wang HW, Pai TW. Machine learning-based methods for prediction of linear B-cell epitopes. Methods Mol Biol 2014. https://doi.org/10.1007/978-1-4939-1115-8_12.
[24] Sela-Culang I, Ashkenazi S, Peters B, Ofran Y. PEASE: Predicting B-cell epitopes utilizing antibody sequence. Bioinformatics 2015. https://doi.org/10.1093/bioinformatics/btu790.
[25] Liu T, Shi K, Li W. Deep learning methods improve linear B-cell epitope prediction. BioData Min 2020. https://doi.org/10.1186/s13040-020-00211-0.
[26] Krawczyk K, Liu X, Baker T, Shi J, Deane CM. Improving B-cell epitope prediction and its application to global antibody-antigen docking. Bioinformatics 2014. https://doi.org/10.1093/bioinformatics/btu190.
[27] Jespersen MC, Mahajan S, Peters B, Nielsen M, Marcatili P. Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes. Front Immunol 2019. https://doi.org/10.3389/fimmu.2019.00298.
[28] Pittala S, Bailey-Kellogg C. Learning context-aware structural representations to predict antigen and antibody binding interfaces. Bioinformatics 2020. https://doi.org/10.1093/bioinformatics/btaa263.
[29] Dai B, Bailey-Kellogg C. Protein interaction interface region prediction by geometric deep learning. Bioinformatics 2021. https://doi.org/10.1093/bioinformatics/btab154.
[30] Kim DG, Choi Y, Kim HS. Epitopes of protein binders are related to the structural flexibility of a target protein surface. J Chem Inf Model 2021. https://doi.org/10.1021/acs.jcim.0c01397.
[31] Wong WK, Robinson SA, Bujotzek A, Georges G, Lewis AP, Shi J, et al. Ab-Ligity: identifying sequence-dissimilar antibodies that bind to the same epitope. MAbs 2021. https://doi.org/10.1080/19420862.2021.1873478.
[32] Brooks BD, Closmore A, Yang J, Holland M, Cairns T, Cohen GH, et al. Characterizing epitope binding regions of entire antibody panels by combining experimental and computational analysis of antibody: antigen binding competition. Molecules 2020. https://doi.org/10.3390/molecules25163659.
[33] Lee JJ, Kim HJ, Yang CS, Kyeong HH, Choi JM, Hwang DE, et al. A high-affinity protein binder that blocks the IL-6/STAT3 signaling pathway effectively suppresses non-small cell lung cancer. Mol Ther 2014. https://doi.org/10.1038/mt.2014.59.
[34] Heu W, Choi JM, Lee JJ, Jeong S, Kim HS. Protein binder for affinity purification of human immunoglobulin antibodies. Anal Chem 2014. https://doi.org/10.1021/ac501158t.
[35] Choi Y, Jeong S, Choi JM, Ndong C, Griswold KE, Bailey-Kellogg C, et al. Computer-guided binding mode identification and affinity improvement of an LRR protein binder without structure determination. PLoS Comput Biol 2020. https://doi.org/10.1371/journal.pcbi.1008150.
[36] Kim HY, Lee JJ, Kim N, Do Heo W, Kim HS. Tracking protein-protein interaction and localization in living cells using a high-affinity molecular binder. Biochem Biophys Res Commun 2016. https://doi.org/10.1016/j.bbrc.2016.01.129.
[37] Hwang DE, Choi JM, Yang CS, Lee JJ, Heu W, Jo EK, Kim HS. Effective suppression of C5a-induced proinflammatory response using anti-human C5a repebody. Biochem Biophys Res Commun 2016. https://doi.org/10.1016/j.bbrc.2016.07.041.
[38] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2017. https://doi.org/10.1038/nprot.2016.169.

[39] Somers W, Stahl M, Seehra JS. 1.9 Å crystal structure of interleukin 6: implications for a novel mode of receptor dimerization and signaling. EMBO J 1997. https://doi.org/10.1093/emboj/16.5.989.
[40] Lensink MF, Wodak SJ. Score_set: A CAPRI benchmark for scoring protein complexes. Proteins Struct Funct Bioinforma 2014. https://doi.org/10.1002/prot.24678.
[41] Radom F, Plückthun A, Paci E. Assessment of ab initio models of protein complexes by molecular dynamics. PLoS Comput Biol 2018. https://doi.org/10.1371/journal.pcbi.1006182.
[42] Jandova Z, Vargiu AV, Bonvin AMJJ. Native or non-native protein-protein docking models? Molecular dynamics to the rescue. J Chem Theory Comput 2021. https://doi.org/10.1021/acs.jctc.1c00336.
[43] de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 2010;5:883–97. https://doi.org/10.1038/nprot.2010.32.
[44] Hua CK, Gacerez AT, Sentman CL, Ackerman ME, Choi Y, Bailey-Kellogg C. Computationally-driven identification of antibody epitopes. Elife 2017. https://doi.org/10.7554/eLife.29023.
[45] Sohn YK, Son S, Choi Y, Hwang DE, Seo HD, Lee JJ, Kim HS. Effective inhibition of C3a-mediated pro-inflammatory response by a human C3a-specific protein binder. Biotechnol Bioeng 2020. https://doi.org/10.1002/bit.27309.
[46] Cook WJ, Choi Y, Gacerez A, Bailey-Kellogg C, Sentman CL. A Chimeric antigen receptor that binds to a conserved site on MICA. ImmunoHorizons 2020. https://doi.org/10.4049/immunohorizons.2000041.
[47] Sela-Culang I, Benhnia MREI, Matho MH, Kaever T, Maybeno M, Schlossman A, et al. Using a combined computational-experimental approach to predict antibody-specific B cell epitopes. Structure 2014. https://doi.org/10.1016/j.str.2014.02.003.
[48] Scally SW, McLeod B, Bosch A, Miura K, Liang Q, Carroll S, et al. Molecular definition of multiple sites of antibody inhibition of malaria transmission-blocking vaccine antigen Pfs25. Nat Commun 2017. https://doi.org/10.1038/s41467-017-01924-3.
[49] Hansen J, Baum A, Pascal KE, Russo V, Giordano S, Wloga E, et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science (80-) 2020. https://doi.org/10.1126/science.abd0827.
[50] Crowe JE. Influenza virus-specific human antibody repertoire studies. J Immunol 2019. https://doi.org/10.4049/jimmunol.1801459.
[51] Šali A. MODELLER A program for protein structure modeling. Comp Protein Model by Satisf Spat Restraints 1993:779–815.
[52] Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006. https://doi.org/10.1110/ps.062416606.
[53] Ponder JW, Case DA. Force fields for protein simulations. Adv Protein Chem 2003. https://doi.org/10.1016/S0065-3233(03)66002-X.
[54] Clark Still W, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990. https://doi.org/10.1021/ja00172a038.
[55] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins Struct Funct Genet 2006. https://doi.org/10.1002/prot.21123.
[56] Brenke R, Hall DR, Chuang GY, Comeau SR, Bohnuud T, Beglov D, et al. Application of asymmetric statistical potentials to antibody-protein docking. Bioinformatics 2012. https://doi.org/10.1093/bioinformatics/bts493.
[57] Waskom M. seaborn: statistical data visualization. J Open Source Softw 2021. https://doi.org/10.21105/joss.03021.
[58] Leem J, Dunbar J, Georges G, Shi J, Deane CM. ABodyBuilder: Automated antibody structure prediction with data–driven accuracy estimation. MAbs 2016. https://doi.org/10.1080/19420862.2016.1205773.