bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Unlocking *de novo* antibody design with generative artificial intelligence

Amir Shanehsazzadeh*, Sharrol Bachas*, Matt McPartlon*, George Kasun, John M. Sutton, Andrea K. Steiger, Richard Shuai, Christa Kohnert, Goran Rakocevic, Jahir M. Gutierrez, Chelsea Chung, Breanna K. Luton, Nicolas Diaz, Simon Levine, Julian Alverio, Bailey Knight, Macey Radach, Alex Morehead, Katherine Bateman, David A. Spencer, Zachary McDargh, Jovan Cejovic, Gaelin Kopec-Belliveau, Robel Haile, Edriss Yassine, Cailen McCloskey, Monica Natividad, Dalton Chapman, Joshua Bennett, Jubair Hossain, Abigail B. Ventura, Gustavo M. Canales, Muttappa Gowda, Kerianne A. Jackson, Jennifer T. Stanton, Marcin Ura, Luka Stojanovic, Engin Yapici, Katherine Moran, Rodante Caguiat, Amber Brown, Shaheed Abdulhaqq, Zheyuan Guo, Lillian R. Klug, Miles Gander, Joshua Meier✉

Absci Corporation, New York (NY) and Vancouver (WA), USA

* Equal contribution
✉ Corresponding author (jmeier@absci.com)

Abstract

Generative artificial intelligence (AI) has the potential to greatly increase the speed, quality and controllability of antibody design. Traditional *de novo* antibody discovery requires time and resource intensive screening of large immune or synthetic libraries. These methods also offer little control over the output sequences, which can result in lead candidates with sub-optimal binding and poor developability attributes. Several groups have introduced models for generative antibody design with promising *in silico* evidence [1–10], however, no such method has demonstrated generative AI-based *de novo* antibody design with experimental validation. Here we use generative deep learning models to *de novo* design antibodies against three distinct targets, in a *zero-shot* fashion, where all designs are the result of a single round of model generations with no follow-up optimization. In particular, we screen over 1 million antibody variants designed for binding to human epidermal growth factor receptor 2 (HER2) using our high-throughput wet lab capabilities. Our models successfully design all CDRs in the heavy chain of the antibody and compute likelihoods that are calibrated with binding. We achieve binding rates of 10.6% and 1.8% for heavy chain CDR3 (HCDR3) and HCDR123 designs respectively, which is four and eleven times higher than HCDR3s and HCDR123s randomly sampled from the Observed Antibody Space (OAS) [11]. We further characterize 421 AI-designed binders using surface plasmon resonance (SPR), finding three that bind tighter than the therapeutic antibody trastuzumab. The binders are highly diverse, have low sequence identity to known antibodies, and adopt variable structural conformations. Additionally, the binders score highly on our previously introduced *Naturalness* metric [12], indicating they are likely to possess desirable developability profiles and low immunogenicity. We open source[1] the HER2 binders and report the measured binding affinities. These results unlock a path to accelerated drug creation for novel therapeutic targets using generative AI and high-throughput experimentation.

[1] https://github.com/AbsciBio/unlocking-de-novo-antibody-design

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Antibodies are a growing class of therapeutic molecules [13] due to their attractive drug-like properties, including high target selectivity and minimal immunogenic effects [14]. Antibody drug development commonly begins with initial lead molecule discovery. Existing approaches for lead discovery typically consist of randomly searching through a massive combinatorial sequence space by screening large libraries of random antibody variants against a target antigen. Techniques such as phage display [15], yeast display [16], immunization coupled with hybridoma screening [17] or B-cell sequencing [18] are typically employed for initial discovery, followed by further molecule development. These methods are time and resource intensive, lack control over the properties of the resulting antibodies, and often produce sub-optimal leads. Applying generative artificial intelligence (AI) to design *de novo* antibodies in a *zero-shot* and controllable fashion, rather than screening and developing lead molecules, could drastically reduce the time and resources necessary for therapeutic antibody development. The application of AI methods to antibody design, and more generally protein therapeutic design, is compelling given the availability of large protein sequence and structure databases that can be used for model training [11, 19–46]. Indeed, recent work [47–51] has shown that models trained on these data could be used for the *de novo* design of certain classes of proteins. These works screen dozens to thousands of protein designs, representing two to four orders of magnitude fewer proteins than are validated in our study. Moreover, no method has yet achieved zero-shot AI-based *de novo* design of antibodies with wet lab validation, despite the immense therapeutic relevance of antibody-based therapeutics which accounted for 30% of FDA approved biologics in 2022 [52].

Here, we follow the definition of *zero-shot* as provided in the GPT-3 [53] and ESM-1v [54] papers. For antibodies, this means designing an antibody to bind to an antigen with no previous demonstrations of binders to said antigen. In other words, the model is used *directly* for design and no further optimization is performed. We also go further to show that our design process is *de novo*, which is traditionally defined as the design of proteins from first principles or more broadly the use of engineering principles to design antibodies from scratch [55, 56]. To demonstrate this, we remove training data containing antibodies known to bind the target or any homolog to the target, which means the model needs to design the antibody *from scratch* or *de novo*. Several works [57–60] have succeeded in optimizing antibodies using supervised learning, though none have validated zero-shot or *de novo* antibody design with experimental validation. Our work focuses on the experimental validation of a generative AI approach. We leave discussion of modeling and ablation studies to possible follow up publications.

Many groups have recognized the potential of zero-shot generative AI to impact antibody design. Several methods have recently emerged, leveraging ideas from language modeling to geometric learning, for the design of antibodies [1–10]. As in our work, these papers focus on the design of the complementary determining regions (CDRs), as these are the key determinants of antibody function and interact directly with the antigen [61–63]. However, no such method has been able to demonstrate *de novo* antibody design in a zero-shot fashion with validation in the lab. Key to our work is the integration of novel generative modeling ideas with high-throughput experimentation capabilities in the wet lab. Recent advancements in DNA synthesis and sequencing, *E. coli* based antibody expression, and fluorescence-activated cell sorting have made it possible to experimentally assess hundreds of thousands of individual designs rapidly and in parallel [64].

Here we demonstrate zero-shot antibody design with extensive wet lab experimentation. As a first step towards fully *de novo* antibody design, we show that all CDRs in the heavy chain of an antibody can be designed with generative AI methods using trastuzumab and its target antigen, HER2, as a model system [65, 66]. All antibodies

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that bind HER2 or homologs of HER2 are removed from the training set. We screen over 1 million unique HCDR3 and HCDR123 variants of trastuzumab and screen for binding to HER2 using our proprietary Activity-specific Cell-Enrichment (ACE) assay [64]. From these designs, we functionally validate 421 binders using SPR. Not only do the designed binders possess sequence novelty compared to those found in the training dataset, they are also highly diverse and dissimilar to anything previously observed in structural antibody databases [22] or massive datasets of known antibodies [11]. According to our previously described *Naturalness* metric [12], the designed binders are likely to be developable and possess favorable immunogenicity characteristics. Furthermore, 3D predicted structures of the *de novo* designed HCDR3s bound to HER2 reveal large conformational variability but discrete spatially conserved side chains when compared to trastuzumab. We show the extensibility of our approach by designing and validating binding molecules to two additional antigens: human vascular endothelial growth factor A (VEGF-A) and the SARS-CoV-2 spike RBD (COVID-19 Omicron variant).

An earlier version of this manuscript in January 2023 focused on the *in silico* design of heavy chain CDR3 (HCDR3). In a revision (March 2023) the model is applied to the design of all three heavy chain CDRs (HCDR123). Additional changes in the revision (March 2023) include extensive characterization of model binding rates (>10% for HCDR3, >1% for HCDR123) and execution of the largest biological baseline study in the field of generative AI-based *de novo* antibody design. Taken together, this work paves the way for rapid progress toward fully *de novo* antibody design using generative AI, which has the potential to revolutionize the availability of therapeutics for patients.

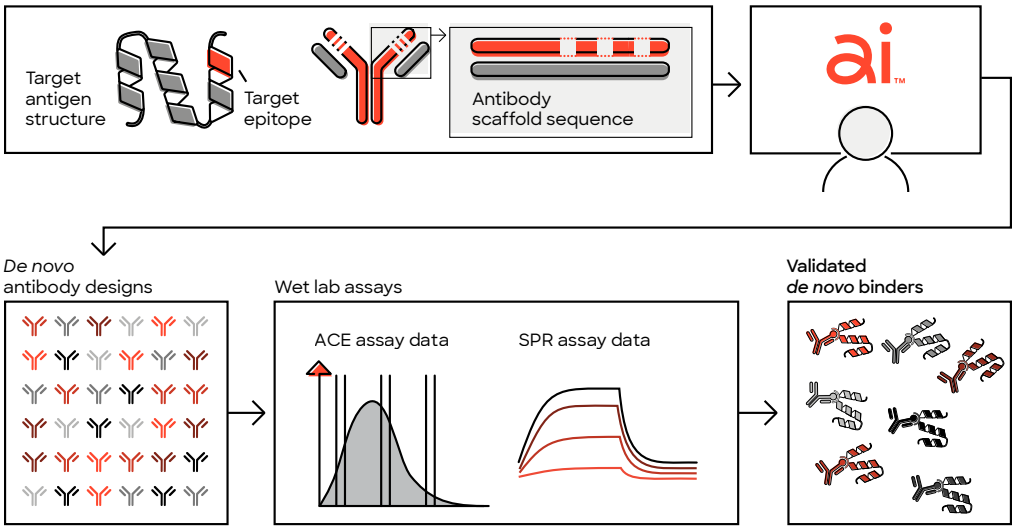

Fig 1. *Zero-shot* generative AI for *de novo* antibody design. Deep learning models trained on antibody-antigen interactions combined with high-throughput wet lab experimentation enable the design of binders to antigens never-before-seen by the model without need for further affinity maturation or lead optimization. Inputs to the model consist of target antigen structure, target epitope region, and antibody framework sequences. None of the CDR sequences are provided to the model as input.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Screening model generated sequences for binding

We leverage our previously described ACE assay (Materials and Methods) [12, 64] to screen massive antibody variant libraries containing hundreds of thousands of members expressed in Fragment antigen-binding (Fab) format. We validate the ACE assay for our *de novo* discovery workflow by sampling sequences for follow-up analysis by SPR (Materials and Methods, Figure S1), an industry standard in binding affinity measurement and detection. We find that the ACE assay is routinely able to classify binders (Materials and Methods) with nearly 95% precision and >95% recall (Table S1, S2, Figure S2). This enables a powerful workflow where a large population of predictions can be initially screened by the ACE assay and the expected binding population can be subsequently screened via SPR to remove false positives and collect high quality binding affinity measurements (Figure 1). The high precision and high recall offered by ACE allows us to measure binding rates and compare them across populations. In any direct comparison shown based on ACE data, sequences from all populations are synthesized in the same library and screened in the same ACE assay.

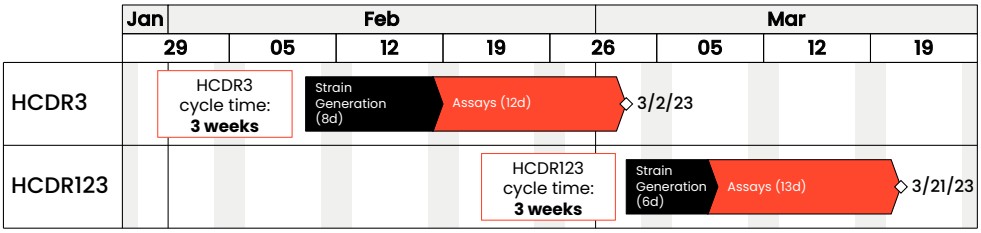

Fig 2. Fast cycle times enable rapid experimentation. We successfully screen >400,000 member antibody libraries in 3 week cycles. All sequences are uniquely specified in the design (e.g. designed by a model or included as a control or baseline). "Strain generation" includes cloning, quality control, and cultivation. "Assays" includes ACE, sequencing, and analysis. A precursor to our screening workflow is DNA synthesis, which is done by a third party vendor and typically takes 7-14 business days.

Zero-shot design of all heavy chain CDRs significantly outperforms biological baselines

Experimental setup. Here, we demonstrate the ability of generative AI models to *de novo* design antibodies targeting specific antigens. To this end, we generate heavy chain CDR3 (HCDR3) and all heavy chain CDR (HCDR123) sequences in a zero-shot fashion. We focus on design of CDRs, key determinants of antibody function, due to their high sequence diversity in immune repertoires and high density of paratope residues [63]. We select trastuzumab, which binds to HER2, as a scaffold antibody to test designed HCDR sequences. We design HCDRs using a model conditioned on the HER2 antigen backbone structure derived from PDB:1N8Z (Chain C) [67] and the trastuzumab framework sequences. None of the CDRs (including light chain CDRs) are provided to the model as inputs.

Furthermore, to show that the model leverages antigen information, we also design HCDRs using a model conditioned on incorrect antigens, specifically rat HER2, HER3, and VEGF, instead of human HER2. The sampled sequences are tested for binding to human Her2. Structural inputs are described in Materials and Methods.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. AI designs achieve high binding rates and outperform biological baselines. Top 1,000 sequences by model likelihood are tested experimentally. * $p < 0.01$, Fisher's exact tests vs. each of the biological baselines and "Wrong Antigens."

Method	**Binding Rate (%)**	
	HCDR3	**HCDR123**
De novo (Human HER2)	**10.6***	**1.8***
Wrong Antigen (Rat HER2)	2.8	0.5
Wrong Antigen (HER3)	2.9	0.2
Wrong Antigen (VEGF)	2.5	0.0
OAS	2.68	0.16
OAS-J	5.25	0.32
SAbDab	3.16	0.06
Permuted sequences	0.33	N/A

Training. We remove any antibody known to bind the target or any homolog of the target (>40% sequence identity or part of the same homologous superfamily) from the training set. In some settings, we instead remove all antibodies from the training set with >40% sequence identity to the wildtype antibody. In all cases, we observe binders. For experiments comparing binding rates, we use the setting where any antibody known to bind the target or any homolog of the target (>40% sequence identity) is removed from the training set.

Inference. We generate sequences using several in-house generative models and hyperparameter settings. We sample sequences from each model independently. Then, we aggregate all designed sequences, compute likelihoods according to each model, and take the average of those scores as the final score used for ranking. Finally, we sample the top k sequences (Materials and Methods) according to this ensembled score and measure the binding rate of the population in the ACE assay (Materials and Methods).

A fraction of each library is also used for ablation studies (results not presented; used for internal benchmarking) so the overall binding rate of the entire libraries (consisting of hundreds of thousands of sequences) may appear lower than the top k binding rates reported here.

As the optimal HCDR lengths cannot be known *a priori*, we sample sequences with HCDR3 lengths corresponding to the distribution of HCDR3 lengths observed in OAS-J (Table S4). This also enables a fair comparison to biological baselines. We always sample HCDR1 and HCDR2 sequences with length 8, as these are the most frequent lengths for HCDR1 (85.4% occurence) and HCDR2 (64.8% occurence) in OAS, as well as the lengths of trastuzumab's HCDR1 and HCDR2, as defined by IMGT [68].

Biological baselines. We compare the binding rates of our generative models to relevant biological baselines, derived by sampling sequences from OAS and SAbDab (Materials and Methods). We screen all unique HCDR3 and HCDR123s in SabDab and randomly sample sequences from OAS. We present the number of binders and sequences for each category in Table S5. We find that the permuted sequences baseline performs worst, which is expected given that randomly shuffling amino acids destroys positional information. The OAS and SAbDab baselines perform similarly, while the OAS-J baseline, which considers sequences from OAS that have the same J-gene as trastuzumab, performs slightly better. This is expected because part of the antibody's HCDR3 is determined by its J-gene and so these sequences have significantly higher homology to

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trastuzumab on average. We find that sequences from our models significantly outperform each of these baselines.

Wet-lab validation. To benchmark our models, we run the largest wet-lab validated baseline study in the field of generative AI-based *de novo* antibody design to date, containing over 100,000 baseline sequences from OAS and SAbDab. We find:

- **AI models outperform biological baselines**: Our *de novo* method achieves a binding rate of 10.6% (HCDR3 design) and 1.8% (HCDR123 design), significantly outperforming the random OAS baseline by 4x and 11x, respectively ($p < 10^{-5}$, Fisher's exact test). Sequences from the model are sampled according to the OAS-J HCDR3 length distribution, enabling a fair comparison (Table 1).
- **Antigen-Specificity**: Performance drops significantly ($p < 10^{-2}$, Fisher's exact test) when using the incorrect antigen as an input, such as rat HER2, HER3, or VEGF (Table 1), indicating the model's use of antigen information for sequence designs. We note that this has not been shown previously for any zero-shot *de novo* antibody design model in the literature. In the case where the wrong antigen is provided to our model, the resulting outputs generally perform similarly to random OAS screening (Table 1) but as more samples are drawn they perform worse than OAS (Table S3).
- **Calibration**: Model likelihood is calibrated with binding, meaning sequences with higher likelihoods are more likely to bind, and therefore sequences can be effectively prioritized in an unsupervised zero-shot manner. We observe this by noting that as more samples are drawn from the model, binding rates tend to decrease (Table 2, S3).

Taken together, these results highlight the effectiveness of Generative AI at designing *de novo* antibodies in a *zero-shot* manner. The computational nature of the approach can enable binder discovery in as few as four weeks (Figure 2).

Table 2. Binding rates of AI-designed *de novo* designs by HCDR3 length. Top k binding rates (i.e. the percentage of the top k sequences that bind) broken down by HCDR3 length for HCDR3 designs ($k = 10, 50, 100$) and HCDR123 designs ($k = 500, 1000$). We also present baseline binding rates among the entire population broken down by HCDR3 length for OAS, OAS-J, and SAbDab.

Population		HCDR3 Length (Binding Rate (%))								
		9	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
HCDR3	$k = 10$	**90**	20	0	0	10	0	**50**	0	0
	$k = 50$	54	**22**	0	6	14	0	24	0	0
	$k = 100$	53	18	1	8	**20**	1	17	0	**2**
	OAS	14.28	5.63	3.45	4.62	2.57	1.41	0.56	0.42	0.52
	OAS-J	15.70	4.86	**5.65**	**10.51**	5.76	**2.96**	1.96	**1.09**	1.38
	SAbDab	9.44	3.63	3.38	5.19	3.73	1.63	0.71	0.42	0
HCDR123	$k = 500$	1.4	**10.4**	**1.2**	0.6	**0.6**	**0.4**	**0.6**	**0.2**	0.2
	$k = 1000$	**2.4**	9.2	0.9	0.4	0.4	**0.4**	0.4	**0.2**	0.1
	OAS	0.24	0.23	0.12	0.31	0.22	0.11	0.14	0.05	0.08
	OAS-J	0.58	0.21	0.52	**0.62**	0.25	0.06	0.32	0.14	**0.22**
	SAbDab	0	0	0.58	0	0	0	0	0	0

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Generative models produce diverse binders

We advance a subset of confirmed binders from an earlier version of these screens for further characterization and confirm HER2 binding for 421 zero-shot AI designs using SPR. Confirmed binders (Figure 3A) show a range of affinity to HER2, with 71 designs exhibiting affinities $<$ 10nM (Figure 3B). Excitingly, three of the zero-shot designs display tighter binding than trastuzumab, with one binding with sub-nanomolar affinity. These high-affinity designs are generated zero-shot from the model without any additional affinity maturation, therefore skipping a typically critical step in the development process of a therapeutic antibody [69]. The ability to generate desirable antibodies that do not need additional optimization could significantly reduce development timelines.

In addition to favorable affinity, the AI model designs have high sequence diversity, both in terms of amino acid length and identity. The verified binders have HCDR3s ranging in length from 11 to 15 amino acids (Figure S3A), compared to the trastuzumab HCDR3 length of 13. The designed sequences are also divergent from the trastuzumab antibody, with edit distances between two and 12 from the trastuzumab sequence (Figure 3C). Average affinity decreases as edit distance increases from the trastuzumab sequence, but interestingly we find designs that still exhibit affinity less than 10nM across all edit distances. We observed one design with an edit distance of nine that exhibits higher affinity than the trastuzumab antibody. Additionally, we found higher diversity in the centers of the HCDR3s, which corresponds to the more diverse D germline gene, compared to the less diverse flanking J and V germline genes [70]. The designs are also sequence diverse from one another, with a mean edit distance of 7.7 ± 2.1 SD (Figure 3D, S3B). Inter-design diversity is noteworthy because it indicates model-generated binders are not converging to shared sequence motifs, as is often seen with traditional antibody screening methods like phage display [71].

Designed binders display sequence novelty

Despite the high sequence diversity of the 421 designed binders, one potential explanation of the model's success is simple reproduction of training examples. This phenomenon has been observed in machine learning models. Indeed, prior methods for generative drug design have been critiqued for generating molecules that are similar to those previously known [72, 73]. Therefore, we compute the minimum distance between the designed binders and all HCDR3s in the model's training and validation sets (Materials and Methods), finding that designed binders are distinct from those observed during training (Figure 4A). We next compute distances to all HCDR3s in the Structural Antibody Database (SAbDab) [22], a database of antibody-antigen complexes, finding that the binder sequences are distant from all antibodies in the database (Figure S4A).

We examined the sequence similarity of the model's outputs to sequences in the Observed Antibody Space (OAS), a database of immune repertoire sequencing studies [11]. We found generated HCDR3s that already exist in the OAS (including those paired with other HCDRs), while others are unique with minimum HCDR3 edit distances between one to five (Figure 4B). Minimum edit distances between all three HCDRs and the HCDRs in OAS are shown in Figure S4B. These results indicate the model is capable of generating biologically relevant yet diverse HCDR3 sequences.

Zero-shot designs are natural

Therapeutic antibody leads that are successful in the drug creation process typically have high affinity and are developable with low immunogenicity. In previous work, we described a language model that can assign a score to antibody sequences indicating the likelihood of finding a sequence in a typical immune repertoire [12]. This metric

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

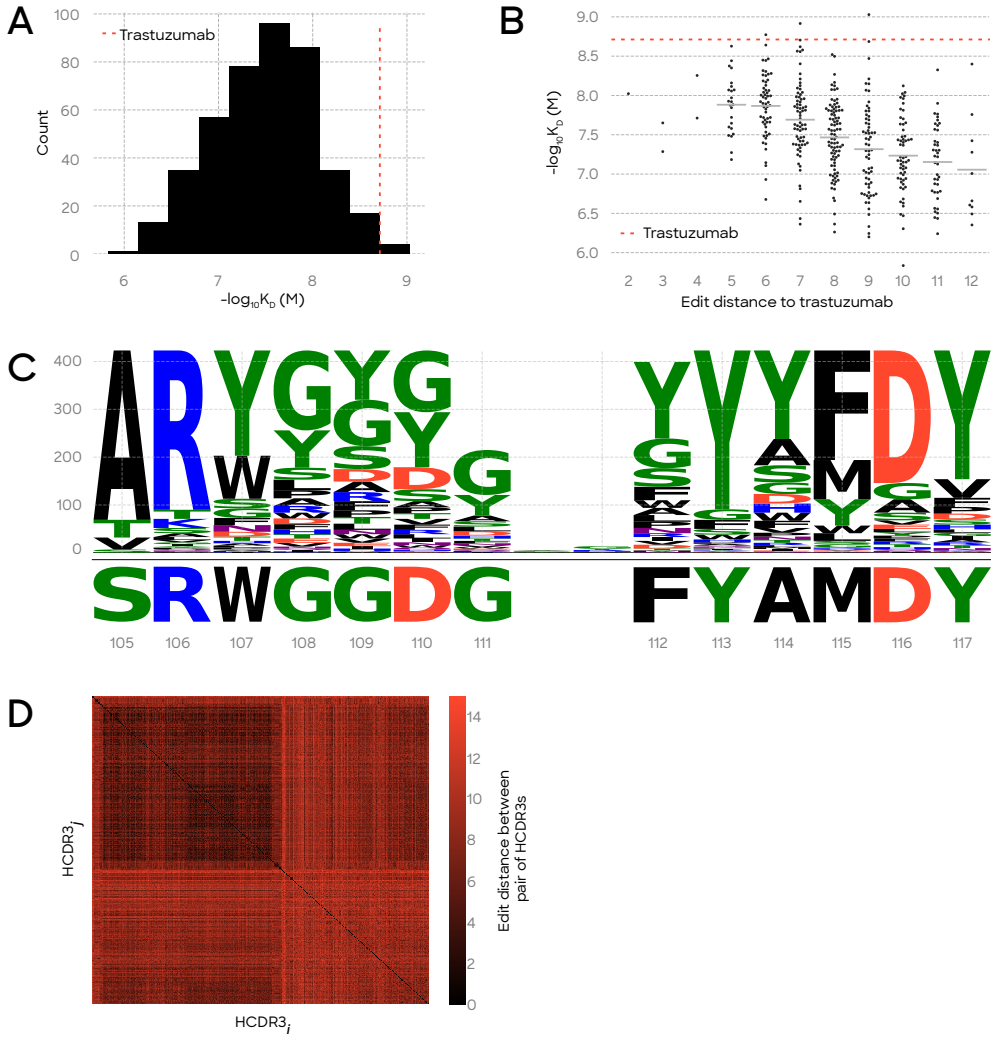

Fig 3. Hundreds of diverse binders created using zero-shot generative AI and validated with SPR. (A) Logo plot of HCDR3s of 421 binding trastuzumab variants. Greater diversity is observed in the centers of the designed HCDR3s. Sequence logo below is the trastuzumab HCDR3 sequence with IMGT [68] numbering shown. **(B)** Binding affinities of AI-generated zero-shot binders. We find 71 designs with comparable affinity (<10 nM) to trastuzumab and 3 with tighter binding. **(C)** Designed variant binding affinities vs. edit distance to trastuzumab. Edit distances range from 2 mutations (84.6% sequence identity) to 12 mutations (7.7% sequence identity). **(D)** Pairwise edit distances between 421 designed binders (minimum of 1, maximum of 15, median of 8, mean of 7.7 ± 2.1 SD). Axes span indices representing the binders.

is referred to as *Naturalness*. A high Naturalness score is associated with favorable antibody developability and immunogenicity. Using the Naturalness scoring model on our designs (Materials and Methods), we find our models can generate sequences with both high affinities and high Naturalness scores in a zero-shot manner, despite not training or sampling based on either metric (Figure 4C). Many designs exhibit Naturalness scores higher than trastuzumab. Figure 4D shows the Naturalness scores for the *de novo* binders as well as several baseline populations. See Table S6 for the mean Naturalness scores across the different populations as well as p-values for the relevant statistical

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

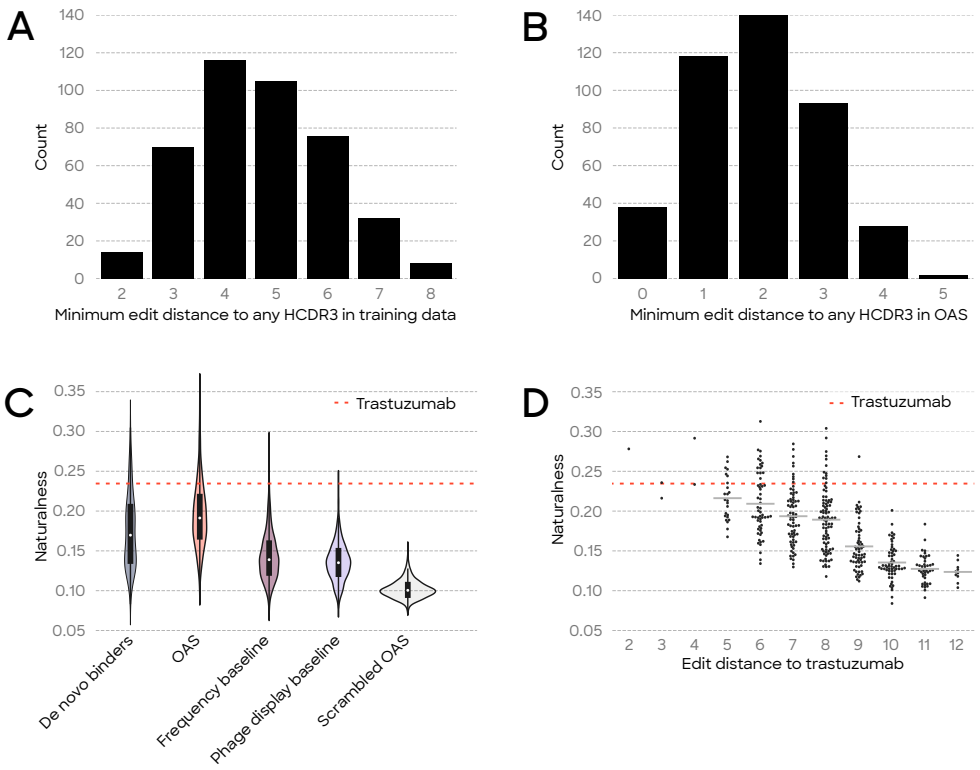

Fig 4. Designed binders are novel and natural. (A) Minimum edit distance of binders to training data HCDR3s (minimum of 2, maximum of 8, median of 5, mean of 4.68 ± 1.34 SD) **(B)** Minimum edit distance of binders to OAS HCDR3s (minimum of 0, maximum of 5, median of 2, mean of 1.91 ± 1.08 SD). 9.3 % (38 out of 421) of the HCDR3 designs are contained in OAS. **(C)** Naturalness scores of designed binders vs. baselines. *De novo* binders are those identified in our study. OAS refers to randomly selected HCDR3s from OAS. Frequency baseline samples amino acids at each position based on positional frequencies observed in OAS for HCDR3. The phage display baseline is a set of HCDR3s sampled from binding and non-binding antibodies from Liu et al [74]. Scrambled OAS are randomly permuted versions of the OAS sequence set. Zero-shot AI designs have significantly higher Naturalness scores on average than the latter three baseline populations ($p < 10^{-50}$) but on average have lower Naturalness scores than trastuzumab and sequences randomly sampled from OAS ($p < 10^{-15}$). Red dashed line is the Naturalness score of trastuzumab. **(D)** Naturalness scores of designed binders vs. edit distance to trastuzumab. Red dashed line is trastuzumab's Naturalness score. Note the presence of a 9-mutation variant with higher Naturalness score than trastuzumab.

comparisons to the *de novo* binders. These results highlight the potential for zero-shot designs to bypass portions of the traditional lead optimization process, potentially saving time and resources in drug development.

Designed binders adopt variable binding mechanisms

We next predict structures for a diverse subset of our *de novo* designed HCDR3 variants to better understand the structural basis of antigen recognition (Materials and Methods). To this end, we built structural models using eight HCDR3 candidates bound to

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HER2 in Fab format. These eight variants are selected based on their edit distance to the trastuzumab HCDR3, binding affinity range (spanning three orders of magnitude) and diversity in length (ranging from 12-15 amino acids) (Table 3, Figure S5). We use the trastuzumab Fab complex with HER2 (PDB:1N8Z) [67] as a starting template for structural modeling. We run local constrained backbone geometry and side chain rotamer optimization followed by relaxation of the complexes to correct global conformational ambiguities, steric clashes, and sub-optimal loop geometry [75]. As a control, we optimize the experimental trastuzumab complex with HER2 using the same protocol for comparison with the optimized HCDR3 structural models. We use the lowest free energy poses of the *de novo* HCDR3 models for structural analyses and comparisons.

Despite the sequence diversity, the eight *de novo* structural models are globally similar to trastuzumab with all-atom HCDR3 RMSDs ranging from 1.9Å − 2.4Å. Minimal structural rearrangements are observed in the unmodified regions of the heavy chain, light chain and epitope residues of the antigen (Figure S6). In select cases, side chains forming contacts with the HCDR3 show slight rotamer differences to account for the presence of longer loops or steric clashes from residues with larger side chains (Figure S7). Alignment of the designed HCDR3 regions with the trastuzumab-HER2 complex reveals a dynamic ensemble of conformations adopted by each HCDR3 (Figure 5). HCDR3 loop structural differences are broad, with RMSDs ranging from 1.1Å − 6.7Å when aligned over all main chain and side chain atoms (Table 3). Even though the *de novo* HCDR3s adopt distinct conformations, there are important positional similarities among all structures (Figure 5). A closer analysis of the spatial orientation of the side chain conformers reveals conservation of identical side chains at five discrete spatial locations. Two of these locations correspond to IMGT residue positions R106 and Y117 in trastuzumab, which are highly conserved in most antibodies [76]. However, there is physiochemical conservation in all structures corresponding to the spatial positions of IMGT residue numbers W107, G109 and Y113 of trastuzumab, which contribute to the paratope of the trastuzumab-HER2 complex [65]. Although conserved spatially, these side chains originate from multiple residue positions, highlighting that conformational flexibility may be required for orienting key paratope residues to form important interactions with HER2.

Although the overall binding region is identical, each designed HCDR3 exhibits distinct binding modes with the epitope. In most cases, novel interactions not observed in the trastuzumab-HER2 complex are formed between the designed HCDR3s and domain IV of HER2 (Figure S7, S8). These interactions are diverse and consist of novel hydrogen bonding interactions, nonpolar interactions, aromatic interactions, and electrostatic interactions formed between each HCDR3 and two distinct surfaces in the HER2 epitope (Figure S7). To further decipher the determinants of binding we calculate the surface area buried by each HCDR3 variant when bound to HER2, which is defined as the binding interface area between paratope and epitope (denoted as Interface in Table 3). In several cases, *de novo* HCDR3 variants show larger binding interface areas than trastuzumab, which would imply novel interactions with the HER2 epitope. Interestingly, no correlation is observed between binding interface area and binding affinity. This suggests that hydrophobic contributions and surface area burial are not key determinants of binding in the designed sequences. Moreover, specific contacts formed between each designed HCDR3 and the epitope are critical to the binding stability of the complex. Furthermore, we calculate the grand average of hydropathy values (GRAVY) [77] of each HCDR3 variant, which defines the collective hydrophobic properties summed over each residue. We compare this to the binding affinities and observe no correlation between affinity and hydrophobicity, which further confirms the hydrophobic effect is not the major determinant of binding for the *de novo* designed HCDR3s (denoted as Hydropathy in Table 3). Combined, these results suggest that the binding affinities of the designed HCDR3s are intrinsic to the sequence design and

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are not driven by a common binding mechanism. The high dependence of binding on sequence attributes agrees with a low probability of designing binders by chance.

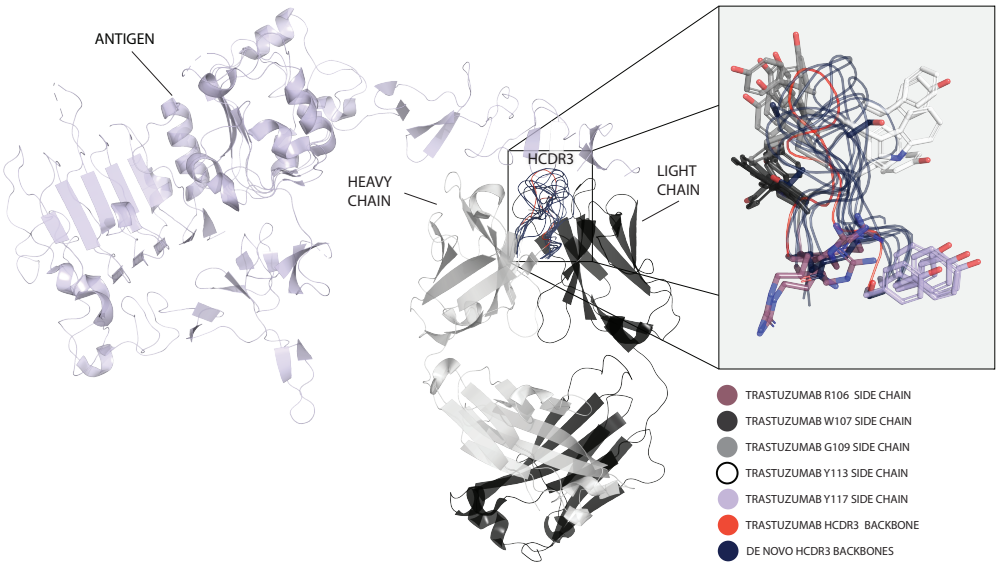

Fig 5. Comparison of trastuzumab-HER2 structure to *de novo* designed binder complexes with HER2. Superimposition of the trastuzumab-HER2 structure with *de novo* designed binder-HER2 complexes shows conformational differences in the HCDR3 backbone. Main chain backbone traces are depicted as ribbons and spatial conserved side chains are shown as sticks. Despite the sequence and length diversity, there are key residues conserved in space, corresponding to the trastuzumab residues W107, G109, and Y113 (IMGT numbering scheme). Residues R106 and Y117 are also conserved which is observed in most HCDR3s.

Table 3. Properties of diverse HCDR3 candidates selected for 3D structural modeling. We select HCDR3 candidates based on affinity, length, and edit distance to trastuzumab. We compute RMSD values over all main chain and side chain atoms from the alignment of HCDR3 residues. All other atoms were excluded from calculations. We calculate grand average of hydropathy values for HCDR3 residues by averaging the hydropathy values of each residue and dividing by sequence length [77].

HCDR3 (Length) [ED]	$-\log_{10}(K_D)$	Interface (Å^2)	RMSD (Å)	Hydropathy
SRWGGDGFYAMDY (13) [0]	8.71	771	0.000	-0.81
ARWGNYYYYMDY (12) [6]	8.77	739	2.435	-1.30
ARYYYGFYYFDY (12) [7]	8.92	819	2.832	-0.73
ARYAGVERPGSFAY (14) [11]	6.24	764	1.107	-0.42
TRYFFNGWYYFDV (13) [9]	9.03	843	1.974	-0.37
AFADSGAYGIWSF (13) [12]	7.00	824	5.738	0.57
ANDIYIQGYDLNR (13) [12]	8.40	833	5.506	-0.80
ARGYSGDWPYETFYV (15) [10]	7.01	863	6.767	-0.76
ARYDYGYYIYVS (12) [10]	8.02	718	3.032	-0.43

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Validation on additional targets

We next conduct a pilot study to demonstrate the applicability of our approach to a broader set of antigens. For these additional targets, we do not pre-screen by the ACE assay. Rather, we sample a small number of sequences and validate binding by SPR. We first successfully design an HCDR3 variant of the therapeutic ranibizumab [78], which binds to human vascular endothelial growth factor A (VEGF-A) (Figure S9, S10). The binder has an affinity of 48.2 nM, as measured by SPR, compared to sub-nanomolar binding of ranibizumab (0.37 nM). Additionally, the designed HCDR3 is highly divergent from ranibizumab, with an edit distance of 13, and novel, with a minimum of 4 mutations separating it from any HCDR3 in OAS.

We design HCDR3 variants of casirivimab [79], conditioned on the Omicron SARS-CoV-2 spike RBD. Casirivimab binds to multiple COVID spike protein variants and, in particular, binds weakly to Omicron. Using SPR, we measure casirivimab affinity to Omicron at $K_D = 240.0$ nM (Figure S11, Table S7). We identify one AI-designed variant that binds with similar affinity to Omicron at $K_D = 179.7$ nM (Table S7). Interestingly, we observe no binding to other spike protein variants for our AI-design, suggesting the potential for controllability of target specificity among homologous antigens. The designed variant has a distinct HCDR3 sequence compared to casirivimab (Figure S12), with an edit distance of 6 and a minimum edit distance of 2 from any HCDR3 in OAS. Additionally, the HCDR3 has at least an edit distance of 4 from any HCDR3 in CoV-AbDab [80].

These binding designs to two additional antigens highlight the extensibility of our zero-shot design approach and indicate the potential for selective antigen controllability with generative AI.

Mult-step CDR design

In an earlier version of this manuscript (January 2023), we reported successful design of multiple HCDRs using a multi-step generative AI design method, which we validated by SPR (Table S8). As of the March 2023 revision, we have deprecated this method (Materials and Methods) internally in favor of our latest *zero-shot* methods, which have now been validated for multiple HCDRs.

Discussion

A particularly difficult aspect of antibody drug creation is the initial step of lead candidate identification due to the labor intensive and uncontrolled nature of traditional screening methods. Generative AI-based *de novo* design has the potential to disrupt these shortcomings of the current drug discovery process. The zero-shot nature of our AI design approach obviates the need for cumbersome library screening to identify binding molecules, generating large time and cost savings. Furthermore, the controllable nature of model-based design allows for the creation of proteins optimized for developability and immunogenicity characteristics, mitigating downstream developability risks. The high binding affinities observed could obviate the need for affinity maturation, representing additional time and cost savings. The approach could be deployed for sophisticated design tasks with high therapeutic relevance such as highly specific epitope targeting.

Here we show important progress for *de novo* antibody design by demonstrating the ability to generate, in a zero-shot fashion, novel antibody variants that confer binding and natural sequence characteristics comparable and, in some cases, superior to the parent antibody. We use our models to generate HCDR3 and HCDR123 designs in a *de novo* fashion, achieving significantly higher binding rates than relevant biological baselines.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our AI-generated sequences are distinct from any observed in the model training set and the vast majority are distinct from the known sequences in the OAS database [11], yet maintain high Naturalness scores, showing the model can design antibody sequences along a biologically feasible manifold. Furthermore, the designed sequences are highly dissimilar from one another, indicating the ability to design a diverse solution set of binding molecules. Structural modeling of a subset of the *de novo* HCDR3 binders reveals high backbone conformational variability, but preservation of important contact positions with the HER2 antigen. Finally, we highlight the generalizability of our approach by deploying these generative design methods to distinct antigens.

Building on the demonstrated progress, future work will expand generative design to enable the *de novo* design of all CDRs and framework regions, further diversifying possible binding solutions. Developing epitope-specificity across multiple antigens for antibody designs could allow for precise interaction with biologically relevant target regions associated with disease mechanisms of action. In addition to advancements on the generative modeling front, the speed and scale of wet lab validation for AI-generated designs will progressively increase as the time and cost of DNA synthesis continue to decline.

Our work represents an important advancement in *in silico* antibody design with the potential to revolutionize the availability of effective therapeutics for patients. Generative AI-designed antibodies will significantly reduce development timelines by generating molecules with desired qualities without the need for further optimization. Additionally, the controllability of AI-designed antibodies will enable the creation of customized molecules for specific disease targets, leading to safer and more efficacious treatments than would be possible by traditional development approaches. Our core platform of generative AI design methods and high-throughput wet lab screening capabilities will continue to drive progress on this front, unlocking new capabilities in the rapidly accelerating field of protein therapeutic design.

Acknowledgments

The authors wish to thank Matthew Weinstock and Alec Jaeger for early discussions; Jens Plassmeier, Mario Sanches, Bradley Emi, Thomas Wrona, Sarah Korman, Zach Jonasson, Joseph Sirosh, Ivana Magovcevic-Liebisch, Dan Rabinovitsj, Daniele Biasci and Victor Greiff for critical review of this manuscript; Joe Kaiser, Jonathan Eads, Kelechi Fletcher, Robert Pfingsten, Chris Rudnicky, Chris Vaillancourt, and Bob Albrecht for providing engineering, MLOps, and DevOps support; Stephanie Yasko, and Marcin Klapczynski for schematics and formatting support; Greg Schiffman, Andreas Busch, and Sean McClain for continual support.

Competing interest statement

The authors are current or former employees, contractors, interns, or executives of Absci Corporation and may hold shares in Absci Corporation. Methods and compositions described in this manuscript are the subject of one or more pending patent applications.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Shuai RW, Ruffolo JA, Gray JJ. Generative language modeling for antibody design. bioRxiv. 2022;doi:10.1101/2021.12.13.472419.

2. Jin W, Wohlwend J, Barzilay R, Jaakkola TS. Iterative Refinement Graph Neural Network for Antibody Sequence-Structure Co-design. arXiv:2110.04624 [q-bio.BM]; 2022.

3. Jin W, Barzilay R, Jaakkola T. Antibody-Antigen Docking and Design via Hierarchical Equivariant Refinement. arXiv:2207.06616 [q-bio.BM]; 2022.

4. Luo S, Su Y, Peng X, Wang S, Peng J, Ma J. Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures. bioRxiv. 2022;doi:10.1101/2022.07.10.499510.

5. Nijkamp E, Ruffolo J, Weinstein EN, Naik N, Madani A. ProGen2: Exploring the Boundaries of Protein Language Models. arXiv:2206.13517 [cs.LG]; 2022.

6. Mahajan SP, Ruffolo JA, Frick R, Gray JJ. Hallucinating structure-conditioned antibody libraries for target-specific binders. bioRxiv. 2022;doi:10.1101/2022.06.06.494991.

7. Kong X, Huang W, Liu Y. Conditional Antibody Design as 3D Equivariant Graph Translation. arXiv:2208.06073 [q-bio.BM]; 2022.

8. Gao K, Wu L, Zhu J, Peng T, Xia Y, He L, et al. Incorporating Pre-training Paradigm for Antibody Sequence-Structure Co-design. arXiv:2211.08406 [q-bio.BM]; 2022.

9. Shi C, Wang C, Lu J, Zhong B, Tang J. Protein Sequence and Structure Co-Design with Equivariant Translation. arXiv:2210.08761 [q-bio.BM]; 2022.

10. Ingraham J, Baranov M, Costello Z, Frappier V, Ismail A, Tie S, et al. Illuminating protein space with a programmable generative model. bioRxiv. 2022;doi:10.1101/2022.12.01.518682.

11. Olsen TH, Boyles F, Deane CM. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. Protein Science. 2022;31(1):141–146.

12. Bachas S, Rakocevic G, Spencer D, Sastry AV, Haile R, Sutton JM, et al. Antibody optimization enabled by artificial intelligence predictions of binding affinity and naturalness. bioRxiv. 2022;doi:10.1101/2022.08.16.504181.

13. Kaplon H, Crescioli S, Chenoweth A, Visweswaraiah J, Reichert JM. Antibodies to watch in 2023. In: Mabs. vol. 15. Taylor & Francis; 2023. p. 2153410.

14. Castelli MS, McGonigle P, Hornby PJ. The pharmacology and therapeutic applications of monoclonal antibodies. Pharmacology research & perspectives. 2019;7(6):e00535.

15. Kretzschmar T, Von Rüden T. Antibody discovery: phage display. Current opinion in biotechnology. 2002;13(6):598–602.

16. Feldhaus MJ, Siegel RW. Yeast display of antibody fragments: a discovery and characterization platform. Journal of immunological methods. 2004;290(1-2):69–80.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Fitzgerald V, Leonard P. Single cell screening approaches for antibody discovery. Methods. 2017;116:34–42.

18. Robinson WH. Sequencing the functional antibody repertoirediagnostic and therapeutic discovery. Nature Reviews Rheumatology. 2015;11(3):171–182.

19. Consortium TU. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research. 2022;51(D1):D523–D531. doi:10.1093/nar/gkac1052.

20. Berman HM. The Protein Data Bank. Nucleic Acids Research. 2000;28(1):235–242. doi:10.1093/nar/28.1.235.

21. Burley SK, Berman HM, Duarte JM, Feng Z, Flatt JW, Hudson BP, et al. Protein Data Bank: A Comprehensive Review of 3D Structure Holdings and Worldwide Utilization by Researchers, Educators, and Students. Biomolecules. 2022;12(10):1425. doi:10.3390/biom12101425.

22. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SAbDab: the structural antibody database. Nucleic acids research. 2014;42(D1):D1140–D1146.

23. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences. 2021;118(15):e2016239118. doi:10.1073/pnas.2016239118.

24. Rao RM, Liu J, Verkuil R, Meier J, Canny J, Abbeel P, et al. MSA Transformer. In: Meila M, Zhang T, editors. Proceedings of the 38th International Conference on Machine Learning. vol. 139 of Proceedings of Machine Learning Research. PMLR; 2021. p. 8844–8856. Available from: `https://proceedings.mlr.press/v139/rao21a.html`.

25. Rao R, Meier J, Sercu T, Ovchinnikov S, Rives A. Transformer protein language models are unsupervised structure learners. In: International Conference on Learning Representations; 2021.Available from: `https://openreview.net/forum?id=fylclEqgvgd`.

26. Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW, editors. Advances in Neural Information Processing Systems. vol. 34. Curran Associates, Inc.; 2021. p. 29287–29303. Available from: `https://proceedings.neurips.cc/paper/2021/file/f51338d736f95dd42427296047067694-Paper.pdf`.

27. Du Y, Meier J, Ma J, Fergus R, Rives A. Energy-based models for atomic-resolution protein conformations. In: International Conference on Learning Representations; 2020.Available from: `https://openreview.net/forum?id=S1e_9xrFvS`.

28. Shanehsazzadeh A, Belanger D, Dohan D. Is Transfer Learning Necessary for Protein Landscape Prediction?; 2020.

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589. doi:10.1038/s41586-021-03819-2.

30. Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, et al. High-resolution ide novo/i structure prediction from primary sequence. 2022;doi:10.1101/2022.07.21.500999.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science. 2022;378(6615):49–56. doi:10.1126/science.add2187.

32. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. 2022;doi:10.1101/2022.04.10.487779.

33. Eguchi RR, Choe CA, Huang PS. Ig-VAE: Generative modeling of protein structure by direct 3D coordinate generation. PLOS Computational Biology. 2022;18(6):e1010271. doi:10.1371/journal.pcbi.1010271.

34. Nijkamp E, Ruffolo J, Weinstein EN, Naik N, Madani A. ProGen2: Exploring the Boundaries of Protein Language Models; 2022.

35. Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. Nature Communications. 2022;13(1). doi:10.1038/s41467-022-32007-7.

36. Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM. Unified rational protein engineering with sequence-based deep representation learning. Nature Methods. 2019;16(12):1315–1322. doi:10.1038/s41592-019-0598-1.

37. Ingraham J, Garg VK, Barzilay R, Jaakkola T. Generative Models for Graph-Based Protein Design. In: Advances in Neural Information Processing Systems; 2019.

38. Rao R, Bhattacharya N, Thomas N, Duan Y, Chen X, Canny J, et al. In: Evaluating Protein Transfer Learning with TAPE. Red Hook, NY, USA: Curran Associates Inc.; 2019.

39. Anand N, Huang P. Generative modeling for protein structures. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc.; 2018.Available from: `https://proceedings.neurips.cc/paper/2018/file/afa299a4d1d8c52e75dd8a24c3ce534f-Paper.pdf`.

40. Hie B, Candido S, Lin Z, Kabeli O, Rao R, Smetanin N, et al. A high-level programming language for generative protein design. 2022;doi:10.1101/2022.12.21.521526.

41. Anishchenko I, Pellock SJ, Chidyausiku TM, Ramelot TA, Ovchinnikov S, Hao J, et al. De novo protein design by deep network hallucination. Nature. 2021;600(7889):547–552. doi:10.1038/s41586-021-04184-w.

42. Lai B, McPartlon M, Xu J. End-to-End deep structure generative model for protein design. 2022;doi:10.1101/2022.07.09.499440.

43. Ogden PJ, Kelsic ED, Sinai S, Church GM. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. Science. 2019;366(6469):1139–1143. doi:10.1126/science.aaw2900.

44. Akbar R, Bashour H, Rawat P, Robert PA, Smorodina E, Cotet TS, et al. Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. mAbs. 2022;14(1). doi:10.1080/19420862.2021.2008790.

45. Akbar R, Robert PA, Weber CR, Widrich M, Frank R, Pavlović M, et al. In silico proof of principle of machine learning-based antibody design at unconstrained scale. mAbs. 2022;14(1). doi:10.1080/19420862.2022.2031482.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

46. Robert PA, Akbar R, Frank R, Pavlović M, Widrich M, Snapkov I, et al. Unconstrained generation of synthetic antibody–antigen structures to guide machine learning methodology for antibody specificity prediction. Nature Computational Science. 2022;2(12):845–865. doi:10.1038/s43588-022-00372-4.

47. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. bioRxiv. 2022;doi:10.1101/2022.12.09.519842.

48. Vázquez Torres S, Leung PJY, Lutz ID, Venkatesh P, Watson JL, Hink F, et al. De novo design of high-affinity protein binders to bioactive helical peptides. bioRxiv. 2022;doi:10.1101/2022.12.10.519862.

49. Goverde C, Wolf B, Khakzad H, Rosset S, Correia BE. De novo protein design by inversion of the AlphaFold structure prediction network. bioRxiv. 2022;doi:10.1101/2022.12.13.520346.

50. Verkuil R, Kabeli O, Du Y, Wicky BI, Milles LF, Dauparas J, et al. Language models generalize beyond natural proteins. bioRxiv. 2022;doi:10.1101/2022.12.21.521521.

51. Eguchi RR, Choe CA, Parekh U, Khalek IS, Ward MD, Vithani N, et al. Deep Generative Design of Epitope-Specific Binding Proteins by Latent Conformation Optimization. bioRxiv. 2022;doi:10.1101/2022.12.22.521698.

52. Mullard A. 2022 FDA approvals; 2023. Available from: `http://dx.doi.org/10.1038/d41573-023-00001-3`.

53. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models Are Few-Shot Learners. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20. Red Hook, NY, USA: Curran Associates Inc.; 2020.

54. Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. In: Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in Neural Information Processing Systems. vol. 34; 2021. p. 29287–29303. Available from: `https://openreview.net/forum?id=uXc42E9ZPFs`.

55. Korendovych IV, DeGrado WF. De novo protein design, a retrospective. Q Rev Biophys. 2020;53(e3):e3.

56. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537(7620):320–327. doi:10.1038/nature19946.

57. Shan S, Luo S, Yang Z, Hong J, Su Y, Ding F, et al. Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. Proceedings of the National Academy of Sciences. 2022;119(11):e2122954119. doi:10.1073/pnas.2122954119.

58. Mason DM, Friedensohn S, Weber CR, Jordi C, Wagner B, Meng SM, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. Nature Biomedical Engineering. 2021;(5):600–612. doi:10.5281/zenodo.4899271.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. Saka K, Kakuzaki T, Metsugi S, Kashiwagi D, Yoshida K, Wada M, et al. Antibody design using LSTM based deep generative model from phage display library for affinity maturation. Scientific Reports. 2021;11:5852. doi:10.1038/s41598-021-85274-7.

60. Makowski EK, Kinnunen PC, Huang J, Wu L, Smith MD, Wang T, et al. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. Nature Communications. 2022;13(1):3788. doi:10.1038/s41467-022-31457-3.

61. Sela-Culang I, Kunik V, Ofran Y. The structural basis of antibody-antigen recognition. Frontiers in immunology. 2013;4:302.

62. Ewert S, Honegger A, Plückthun A. Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. Methods. 2004;34(2):184–199.

63. Akbar R, Robert PA, Pavlović M, Jeliazkov JR, Snapkov I, Slabodkin A, et al. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. Cell Reports. 2021;34(11):108856. doi:10.1016/j.celrep.2021.108856.

64. Liu J. Activity-specific cell enrichment; Patent Publication No. WO 2021/146626, 22.07.2021.

65. Bostrom J, Yu SF, Kan D, Appleton BA, Lee CV, Billeci K, et al. Variants of the Antibody Herceptin That Interact with HER2 and VEGF at the Antigen Binding Site. Science. 2009;323(5921):1610–1614.

66. Iqbal N, Iqbal N. Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications. Molecular Biology International. 2014;2014:852748.

67. Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature. 2003;421(6924):756–760. doi:10.1038/nature01392.

68. Lefranc MP, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, et al. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Developmental & Comparative Immunology. 2005;29(3):185–203.

69. Lu RM, Hwang YC, Liu IJ, Lee CC, Tsai HZ, Li HJ, et al. Development of therapeutic antibodies for the treatment of diseases. Journal of biomedical science. 2020;27(1):1–30.

70. Briney BS, Jr JEC. Secondary mechanisms of diversification in the human antibody repertoire. Frontiers in Immunology. 2013;4. doi:10.3389/fimmu.2013.00042.

71. Smith GP, Petrenko VA. Phage display. Chemical reviews. 1997;97(2):391–410.

72. Lowe D. Has AI Discovered a Drug Now? Guess;. Available from: `https://www.science.org/content/blog-post/has-ai-discovered-drug-now-guess`.

73. Walters P. Dissecting the Hype With Cheminformatics;. Available from: `http://practicalcheminformatics.blogspot.com/2019/09/dissecting-hype-with-cheminformatics.html`.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

74. Liu G, Zeng H, Mueller J, Carter B, Wang Z, Schilz J, et al. Antibody complementarity determining region design using high-capacity machine learning. Bioinformatics. 2019;36(7):2126–2133. doi:10.1093/bioinformatics/btz895.

75. Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nature Methods. 2020;17(7):665–680. doi:10.1038/s41592-020-0848-2.

76. Lee CV, Liang WC, Dennis MS, Eigenbrot C, Sidhu SS, Fuh G. High-affinity Human Antibodies from Phage-displayed Synthetic Fab Libraries with a Single Framework Scaffold. Journal of Molecular Biology. 2004;340(5):1073–1093. doi:10.1016/j.jmb.2004.05.051.

77. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 1982;157(1):105–132. doi:10.1016/0022-2836(82)90515-0.

78. Papadopoulos N, Martin J, Ruan Q, Rafique A, Rosconi MP, Shi E, et al. Binding and neutralization of vascular endothelial growth factor (VEGF) and related ligands by VEGF Trap, ranibizumab and bevacizumab. Angiogenesis. 2012;15(2):171–185. doi:10.1007/s10456-011-9249-6.

79. Razonable RR, Pawlowski C, O'Horo JC, Arndt LL, Arndt R, Bierle DM, et al. Casirivimab–Imdevimab treatment is associated with reduced rates of hospitalization among high-risk patients with mild to moderate coronavirus disease-19. EClinicalMedicine. 2021;40:101102. doi:10.1016/j.eclinm.2021.101102.

80. Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the Coronavirus Antibody Database. Bioinformatics. 2021;37(5):734–735. doi:10.1093/bioinformatics/btaa739.

81. Diwanji D, Trenker R, Thaker TM, Wang F, Agard DA, Verba KA, et al. Structures of the HER2–HER3–NRG1 complex reveal a dynamic dimer interface. Nature. 2021;600(7888):339–343. doi:10.1038/s41586-021-04084-z.

82. Chen Y, Wiesmann C, Fuh G, Li B, Christinger HW, McKay P, et al. Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured fab in complex with antigen 1 1Edited by I. A. Wilson. Journal of Molecular Biology. 1999;293(4):865–881. doi:10.1006/jmbi.1999.3192.

83. Fisher RA. On the Interpretation of 2 from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society. 1922;85(1):87. doi:10.2307/2340521.

84. Abhinandan KR, Martin ACR. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. Molecular Immunology. 2008;45(14):3832–3839.

85. Nakamura Y IT Gojobori T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Research. 2000;28(1):292.

86. Mago T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957–2963.

87. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88. Andrews S. FastQC. A quality control tool for high throughput sequence data; 2010. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, https://www.bibsonomy.org/bibtex/2b6052877491828ab53d3449be9b293b3/ozborn.

89. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–3048. doi:10.1093/bioinformatics/btw354.

90. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

91. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8; 2015.

92. Emsley P, Cowtan K. iCoot/i: model-building tools for molecular graphics. Acta Crystallographica Section D Biological Crystallography. 2004;60(12):2126–2132. doi:10.1107/s0907444904019158.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

Biological Baselines

We use the Observed Antibody Space (OAS) [11] and the Structural Antibody Database (SAbDab) [22] to generate sets of biologically relevant sequences for comparison to our generative models.

For heavy chain CDR3s (HCDR3s): The "Random OAS" baseline is constructed by randomly sampling 50,000 unique HCDR3s from OAS with the only condition being that they have a length between 9 and 17 amino acid residues (for parity with our model generations). We similarly construct the "Random OAS-J" baseline by randomly sampling 10,000 unique HCDR3s from OAS from antibodies that have the same J-gene as trastuzumab, while also imposing the same length constraint as the OAS baseline. For the "SAbDab" baseline we include all 2,395 unique HCDR3s from SAbDab that have lengths between 9 and 17 residues and which do not belong to trastuzumab or its variants. For the "Permuted Sequences" baseline we randomly sample a subset of 5,000 HCDR3s from the "Random OAS" baseline and randomly shuffle each sequence's amino acids to destroy positional information.

For all three heavy chain CDRs (HCDR123s): We take an analogous approach to the HCDR3 baselines, except we restrict to sequences with HCDR1 and HCDR2 each of length 8. We sample 50,000 sequences for "Random OAS" baseline, 10,000 sequences for the "Random OAS-J" baseline, and all 1,572 unique HCDR123s from SAbDab fitting the length criteria.

Labeling Binders with ACE

To determine the success of the ACE assay we include several thousand controls (SPR-validated binders and non-binders) in the libraries. The Binary ACE assay (bACE) produces enrichment scores based on proportional abundances in the specified FACS gates. The P_1 and P_2 enrichment scores are predictive of binding (Figure S2) based on their separation of the binding and non-binding controls.

To label screened sequences as binders we use a threshold on the median P_1 enrichment score (across three replicates, R_1, R_2, R_3) and a separate threshold on the minimum P2 enrichment score (across the same three replicates as P_1). Specifically, given thresholds t_1 and t_2 we call a sequence a binder if:

$$\text{Median}(P_1R_1, P_1R_2, P_1R_3) > t_1 \quad \text{OR} \quad \min(P_2R_1, P_2R_2, P_2R_3) > t_2.$$

Otherwise, we label the sequence a non-binder.

We determine the thresholds t_1, t_2 using a grid-search aimed at maximizing F1-score on the controls included in the library. For the heavy chain CDR3 (HCDR3) library we find that $t_1 = 3.51, t_2 = 8.89$ achieve the highest F1-score. For the all three heavy chain (HCDR123) library we find that $t_1 = 6.87, t_2 = 3.05$ achieve the highest F1-score.

Model Structural Inputs

As input to the models we provide an antigen structure and specify an epitope. For the *de novo* setting we provide the structure of human HER2 from PDB:1N8Z (Chain C) [67] and specify the trastuzumab epitope. In this setting, we successfully design using versions of the antigen structure that contain variable amounts of noise or have been relaxed with Rosetta[2].

To show the model's dependence on the antigen information, we attempt design with three incorrect antigens, namely rat HER2, HER3, and VEGF. For rat HER2, we use

[2]https://www.rosettacommons.org/software

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the structure from PDB:1N8Y (Chain A) [67]. For HER3, we use the structure from PDB:7MN8 (Chain A) [81]. For VEGF, we use the structure from PDB:1CZ8 (Chains A, B) [82]. For rat HER2 and HER3, we specify an epitope based on sequential/structural homology to the trastuzumab epitope of human HER2. For VEGF, we specify the ranibizumab [78] epitope.

Binding Rate of Top k Sequences

To determine the top k sequences we sample k sequences according to model likelihood and the OAS length distribution (Table S4). Specifically:

- Define $\text{Top}_\ell(k_\ell)$ as the number of binders amongst the top k_ℓ sequences, according to model likelihood, with HCDR3 lengths equal to ℓ.
- Let f_ℓ be the frequency of length ℓ HCDR3s in OAS-J for $\ell \in \{9, 10, ..., 17\}$. Then we take $(k_9, k_{10}, ...k_{17})$ such that $\sum_{\ell=9}^{17} k_\ell = k$ and the quantity $\sum_{\ell=9}^{17} |k_\ell - k \cdot f_\ell|$ is minimized.
- Let $\text{Top}(k)$ binding rate $= \sum_{\ell=9}^{17} \text{Top}_\ell(k_\ell)$.

Comparing Binding Rates

For comparing binding rates between two populations we use Fisher's exact test [83]. Specifically, if population 1 consists of b_1 binders and n_1 non-binders and population 2 consists of b_2 binders and n_2 non-binders then:

- The binding rates for population 1 and population 2 are given by $r_1 = \frac{b_1}{b_1+n_1}$ and $r_2 = \frac{b_2}{b_2+n_2}$, respectively.
- The ratio of population 1's binding rate to population 2's binding rate is $\frac{r_1}{r_2}$.
- The p-value (from Fisher's exact test) corresponding to the binding rates of population 1 and 2 is
$$p = \frac{\binom{b_1+n_1}{b_1}\binom{b_2+n_2}{b_2}}{\binom{b_1+n_1+b_2+n_2}{b_1+b_2}}.$$

Naturalness Score

The *Naturalness* score used in this study is computed using the pre-trained antibody language model introduced in [12]. It is based on the pseudo-perplexity of the extended CDRs (defined by a union of the IMGT and Martin definitions [68, 84]) of an antibody heavy chain under the language model. This metric is shown to be predictive of desirable therapeutic properties such as developability and lack of immunogenicity.

For HCDR3 variants, Naturalness scores are computed over a grafting of the HCDR3 into the trastuzumab scaffold. In addition to computing Naturalness scores for our *de novo* binders, we include several baselines:

1. OAS: Consists of 1,000 HCDR3s randomly sampled from OAS. Antibody heavy chain sequences are sampled and the HCDR3s are extracted. We expect sequences from OAS to have high Naturalness scores, given that the Naturalness model is pretrained on OAS, so we treat this as a positive control.
2. Frequency baseline: These are 1,000 sequences generated by randomly sampling from a length-conditioned frequency distribution of amino acids in OAS. We compute $P_L(\ell)$, the probability that an HCDR3 in OAS has length ℓ, and then compute

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the probability of sampling a particular sequence with length ℓ, using an independent factorization based on amino acid frequencies at each position:

$$\mathbb{P}_H(s = a_1 a_2 \cdots a_l | L = \ell) = \mathbb{P}(s_1 = a_1 | L = \ell)\mathbb{P}(s_2 = a_2 | L = \ell) \cdots \mathbb{P}(s_\ell = a_\ell | L = \ell).$$

We sample 1,000 lengths $\ell_1, \ell_2, ..., \ell_{1000} \sim P_L$ and then sample 1,000 sequences according to the sampled lengths: $s^i \sim P_{H|L=\ell_i}$. For this baseline, we expect to see a basic level of Naturalness scores since the statistics underlying OAS encode information about biologically viable antibodies. However, despite the distribution of each amino acid position independently matching the OAS distribution, the sampling approach disregards positional dependencies between groups of amino acids, so we expect lower Naturalness scores than those in the OAS baseline.

3. Phage display baseline: We randomly sample 1,000 HCDR3s from the first round of a phage display panning [74]. Antibody heavy chain sequences are sampled and the HCDR3s are extracted. Note that this collection of sampled antibodies consists of both non-binders and binders.

4. Scrambled OAS: This consists of permuted versions of the 1,000 HCDR3s in the OAS control. For each such HCDR3, we permute its sequence 5 different times, compute Naturalness score using the permuted HCDR3, and report the average across the 5 permutations. The motivation for this as a negative control is that permuting a protein sequence destroys positional information. Lower Naturalness scores of this baseline compared to the first OAS baseline implies that the Naturalness model is able to capture positional information, and is not just considering amino acid composition.

We compare the Naturalness scores of our *de novo* designs to these controls using two-sample t-tests ($H_0 : \mu_1 = \mu_2, H_a : \mu_1 \neq \mu_2$) and compare to trastuzumab using one-sample t-tests with trastuzumab's Naturalness score as the population mean ($H_0 : \mu_1 = \mu, H_a : \mu_1 \neq \mu$).

Cloning

Antibody variants are cloned and expressed in Fab format. To produce ACE assay and SPR datasets, DNA variants of HCDR3 alone or spanning HCDR1 to HCDR3 are purchased as single-stranded DNA (ssDNA) oligo pools (Twist Bioscience). We spot check selected binders by re-purchasing as double stranded DNA eBlocks (Integrated DNA Technologies) or ssDNA oligo pools. Codons are randomly selected from the two most common in *E. coli* B strain [85] for each residue.

Amplification of the ssDNA oligo pools is carried out by PCR according to Twist Bioscience's recommendations, except Q5 high fidelity DNA polymerase (New England Biolabs) is used in place of KAPA polymerase. Briefly, 25 μL reactions consist of 1x Q5 Mastermix, 0.3 μM each of forward and reverse primers, and 10 ng oligo pool. Reactions are initially denatured for 3 min at 95°C, followed by 13 cycles of: 95°C for 20 s; 66°C for 20 s; 72°C for 15 s; and a final extension of 72°C for 1 min. DNA amplification is confirmed by agarose gel electrophoresis, and amplified DNA is subsequently purified (DNA Clean and Concentrate Kit, Zymo Research).

To generate linearized vector, a two-step PCR is carried out to split our plasmid vector carrying Fab format trastuzumab into two fragments in a manner that provides cloning overlaps of approximately 25 nucleotides (nt) on the 5' and 3' ends of the amplified ssDNA oligo pool libraries, or 40 nt on the 5' and 3' ends of IDT eBlocks. Vector linearization reactions are digested with DpnI (New England Bioloabs) and purified from a 0.8% agarose gel using the Gel DNA Recovery Kit (Zymo Research) to eliminate

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

parental vector carry through. Cloning reactions consist of 50 fmol of each purified vector fragment, either 100 fmol PCR-amplified ssDNA oligo pool or 10 pmol eBlock library inserts and 1x final concentration NEBuilder HiFi DNA Assembly (New England Biolabs). Reactions are incubated at 50°C for 25 min using eBlocks or two hours using PCR-amplified oligo pools. Assemblies are subsequently purified using the DNA Clean and Concentrate Kit (Zymo Research). DNA concentrations are measured using a NanoDrop OneC (Thermo Scientific).

For HDLs, Transformax EPI300 (Lucigen) *E. coli* is transformed using the MicroPulser Electroporator (BioRad) with the purified assembly reactions and recovered in 1000 μL of SOC medium cultivated at 30°C for 1 hour. The cell culture is then grown in 20 mL of Teknova LB Broth with 50 μg/mL Kanamycin at 30°C and 80 % humidity with 270 rpm shaking for 18 hours. Plasmids are extracted (Plasmid Midi Kit, Zymo Research) and submitted for QC sequencing. Electrocompetent SoluPro^{TM} host strain is transformed with 20 ng of DNA and recovered in 500 μL of SOC medium cultivated at 30°C for 1 hour.

For LDLs, Absci SoluPro^{TM} host strain is transformed with the purified assembly reactions and grown overnight at 30°C on agar plates containing 50 μg/ml kanamycin and 1 % glucose. Colonies are picked for QC analysis prior to cultivation for induction.

QC Analysis

Quality of high diversity variant libraries is assessed by deep sequencing. Briefly, library plasmid pools are amplified by PCR across the region of interest and sequenced with 2x150 or 2x300 nt reads using the Illumina MiSeq platform with 20 % PhiX. The PCR reaction uses 10 nM primer concentration, Q5 2x master mix (New England Biolabs) and 1 ng of input DNA diluted in H_2O. Reactions are initially denatured at 98°C for 3 min; followed by 30 cycles of 98°C for 10 s, 59°C for 30 s, 72°C for 15 s; with a final extension of 72°C for 2 min. Sequencing results are analyzed for distribution of mutations, variant representation, library complexity and recovery of expected sequences. Metrics include coefficient of variation of sequence representation, read share of top 1 % most prevalent sequences and percentage of designed library sequences observed within the library. Quality of low diversity variant libraries is assessed by performing rolling circle amplification (Equiphi29, Thermo Fisher Scientific) on 24 colonies and sequencing using the Illumina DNA Prep, Tagmentation Kit (Illumina Inc.). Each colony is analyzed for mutations from reference sequence, presence of multiple variants, misassembly, and matching to a library sequence (Geneious Prime).

Antibody Expression in SoluPro^{TM} *E. coli* B Strain

After recovery in SOC medium, HDLs are grown in 50 mL of Teknova LB Broth with 50 μg/mL Kanamycin at 30°C and 80 % humidity with 270 rpm shaking for 24 hours. After 24 hours, the pre-culture is diluted to OD600 = 1 in 100 mL induction base medium (IBM) (4.5 g/L Potassium Phosphate monobasic, 13.8 g/L Ammonium Sulfate, 20.5 g/L yeast extract, 20.5 g/L glycerol, 1.95 g/L Citric Acid) containing inducers and supplements (250 μM Arabinose, 50 μg/mL Kanamycin, 8 mM Magnesium Sulfate, 1 mM Propionate, 1X Korz trace metals) and grown for 16 hours in a 500 mL baffled flask at 26°C and 80 % humidity with 270 rpm shaking. At the end of the 16 hours, 250 μL aliquots adjusted to 20 % v/v glycerol are stored at -80°C.

After transformation and QC of LDLs, individual colonies are picked into deep well plates containing 400 μL of Teknova LB Broth 50 μg/mL Kanamycin and incubated at 30°C and 80 % humidity with 1000 rpm shaking for 24 hours. At the end of the 24 hours, 150 μL samples are centrifuged (3300 g, 7 min), supernatant decanted from the pre-culture plate, and cell pellets sent for sequence analysis. 80 μL of the pre-culture is

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transferred to 400 μL of IBM containing inducers and supplements as described above. Culture is grown for 16 hours at 26°C and 80 % humidity with 270 rpm shaking. After 16 hours, 150 μL samples are taken and centrifuged (3300 g, 7 min) into pellets with supernatant decanting prior to being stored at -80°C.

Activity-specific Cell-Enrichment (ACE) Assay

Cell Preparation

High-throughput screening of antigen-specific Fab-expressing cells is adapted from the approach described in [12,64]. For staining, thawed glycerol stocks from induced cultures are transferred to 0.7 ml matrix tubes (500 μL, OD600 = 2), centrifuged (4000 g, 5 min), and resulting pelleted cells are washed three times with PBS (pH 7.4, 1 mM EDTA). Washed cells are thoroughly resuspended in 250 μL of phosphate buffer (32 mM, pH 7.4) by pipetting prior to fixation by the addition of 250 μL of 0.6 % paraformaldehyde and 0.04 % glutaraldehyde in phosphate buffer (32 mM, pH 7.4). After 40 min incubation on ice, samples are centrifuged (4000 g, 5 min) and pellets are washed three times with PBS (pH 7.4, 1 mM EDTA), resuspended in permeabilization buffer (20 mM Tris, 50 mM glucose, 10 mM EDTA, 5 μg/mL rLysozyme), and incubated for 8 min on ice. Fixed and permeabilized cells are then centrifuged (4000 g, 5 min) and washed three times with staining buffer (Perkin Elmer AlphaLISA immunoassay buffer, 25 mM HEPES, 0.1 % casein, 1 mg/mL dextran-500, 0.5 % Triton X-100, 0.05 % Kathon).

Staining

Prior to library staining, the HER2 probe is titrated against the reference strain to determine the 75 % effective concentration (EC_{75}). Following cell preparation, the library is resuspended in 500 ţL staining buffer containing 100 nM either His/Avi tagged human HER2 (Acro Biosystems) conjugated to 50 nM streptavidin-AF647 (Invitrogen) or tag-free human HER2 (Acro Biosystems) directly conjugated to AF647 via free amines. Libraries are incubated with the probe overnight (16 h) with end-to-end rotation at 4°C, centrifuged (4000 g, 5 min), and pellets are washed three times with PBS. Pellets are then resuspended in 500 μL of staining buffer containing 26.5 nM anti-kappa light chain:BV421 (BioLegend) and incubated for 2 hours with end-to-end rotation at 4°C prior to centrifugation (4000 g, 5 min), three washes with PBS and resuspension in 200 ţL of PBS for sorting.

Sorting

Libraries are sorted by one of two methods based on binding: the previously described ACE Assay designed to give quantitative affinity readouts [12, 64] or a binary version of the ACE Assay. For either method, libraries are sorted on FACSymphony S6 (BD Biosciences) instruments. Immediately prior to sorting, 50 μL of stained sample is transferred to a flow tube containing 1 mL PBS + 3 μL propidium iodide. Aggregates, debris, and impermeable cells are removed with singlets, size, and PI+ parent gating, respectively. Cells are then gated to include only those with kappa light chain expression (BV421). For the quantitative ACE Assay, collection gates are drawn to sample across the log range of binding signal. The far right gate is set to collect the brightest 0.1 % of the library and the far left gate is set to collect at the low end of the positive binding signal based on stained control strains. Four additional gates of the same width are then distributed in between, with each set to be approximately half the gMFI of the gate to the right. For the binary version of the ACE Assay, a total of three collection gates are set to sample at the high end of the binding range (top 0.1-2.5%, depending on overall library positivity), the remaining positive binding signal events, and a negative gate

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

containing the events with no binding signal. Libraries are sorted simultaneously on up to four instruments with photomultipliers adjusted to normalize fluorescence intensity, and the collected events are processed independently as technical replicates.

Next-generation Sequencing

Sorted Material Sample Preparation

Sample preparation for sequencing follows the same protocol for both the previously described ACE Assay and the binary version of the ACE Assay. Cell material from sorted gates is collected in a diluted PBS mixture (VWR), in 1.5 mL tubes (Eppendorf). A sample of the unsorted library material is also processed for QC and ACE Assay metric calculations. Post-sort samples are centrifuged (3,800 g) and tube volume is normalized to 20 μl. Amplicons encompassing the HCDR3 or VH region are generated by PCR. The reaction used 10 nM primer concentration, Q5 2x master mix (New England Biolabs) and 20 μl of sorted cell material input suspended in diluted PBS (VWR). Reactions are initially denatured at 98°C for 3 min, followed by 30 cycles of 98°C for 10 s; 59°C for 30 s; 72°C for 15 s; with a final extension of 72°C for 2 min. After amplification, samples are cleaned enzymatically using ExoSAP-IT (Applied Biosystems). Resulting DNA samples are quantified by Qubit fluorometer (Invitrogen), prepped for sequencing with the ThruPLEX DNA-Seq Kit (Takara Bio), normalized and pooled. Pool size is verified via Tapestation 1000 HS and is sequenced on an Illumina NextSeq 1000 P2 (2x150 nt or 2x300nt) with 20 % PhiX.

ACE Assay Analysis

In order to produce quantitative binding scores from reads, the following processing and quality control steps are performed:

1. Paired-end reads are merged using FLASH2 [86] with the maximum allowed overlap set according to the amplicon size and sequencing reads length (150 bases for all the libraries described in this manuscript).

2. Primers are removed from both ends of the merged read using the cutadapt tool [87], and reads are discarded where primers are not detected.

3. Reads are aggregated across all FACS sorting gates and then discarded if (1) the mean base quality is below 20, or (2) a sequence (in DNA space) is seen in fewer than 10 reads across all gates.

4. FastQC [88] and MultiQC [89] are used to generate sequencing quality control metrics.

5. For each gate, the prevalence of each sequence (read count relative to the total number of reads from all sequences in that gate) is normalized to 1 million counts.

6. The binding score (ACE Assay score) is assigned to each unique DNA sequence by taking a weighted average of the normalized counts across the sorting gates. For all experiments, weights are assigned linearly using an integer scale: the gate capturing the lowest fluorescence signal is assigned a weight of 1, the next lowest gate is assigned a weight of 2, etc.

7. Any detected sequence which is not present in the originally designed and synthesized library is dropped.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8. ACE Assay scores are averaged across independent FACS sorts, dropping sequences for which the standard deviation of replicate measurements is greater than 1.25. An amino acid variant is retained only if we collected at least three independent QC-passing observations between synonymous DNA variants and replicate FACS sorts.

Binary ACE Assay Analysis

Enrichment scores are calculated for individual variants screened by a binary version of the ACE Assay using the following procedure:

1. Paired-end reads are merged using Fastp [90] with quality filtering and base correction in merged regions enabled.
2. Primers are removed from both ends of the merged read using the cutadapt tool [87], and reads are discarded where primers are not detected.
3. Unique sequences are tallied to provide raw counts of each variant observed in each sample. Sequences that did not match a designed sequence in the library are discarded.
4. For each sample, proportional abundances are calculated for each variant. Enrichment scores are calculated by dividing the proportional abundance of each variant in a gate by its proportional abundance in the unsorted library sample.

Surface Plasmon Resonance (SPR)

Sample Preparation

Post induction samples are transferred to 96-well plates (Greiner Bio-One), pelleted and lysed in 50 μL lysis buffer (1X BugBuster protein extraction reagent containing 0.01 KU Benzonase Nuclease and 1X Protease inhibitor cocktail). Plates are incubated for 15-20 min at 30°C then centrifuged to remove insoluble debris. After lysis, samples are adjusted with 200 μL SPR running buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.01 % w/v Tween-20, 0.5 mg/mL BSA) to a final volume of 260 μL and filtered into 96-well plates. Lysed samples are then transferred from 96-well plates to 384-well plates for high-throughput SPR using a Hamilton STAR automated liquid handler. Colonies are prepared in two sets of independent replicates prior to lysis and each replicate is measured in two separate experimental runs. In some instances, single replicates are used, as indicated.

SPR

High-throughput SPR experiments are conducted on a microfluidic Carterra LSA SPR instrument using SPR running buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.01 % w/v Tween-20, 0.5 mg/mL BSA) and SPR wash buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.01 % w/v Tween-20). Carterra LSA SAD200M chips are pre-functionalized with 20 μg/mL biotinylated antibody capture reagent for 600 s prior to conducting experiments. Lysed samples in 384-well blocks are immobilized onto chip surfaces for 600 s followed by a 60 s washout step for baseline stabilization. Antigen binding is conducted using the non-regeneration kinetics method with a 300 s association phase followed by a 900 s dissociation phase. For analyte injections, six leading blanks are introduced to create a consistent baseline prior to monitoring antigen binding kinetics. After the leading blanks, five concentrations of HER2 extracellular domain

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

antigen (ACRO Biosystems, prepared in three-fold serial dilution from a starting concentration of 500 nM), are injected into the instrument and the time series response was recorded. In most experiments, measurements on individual DNA variants are repeated four times. Typically each experiment run consists of two complete measurement cycles (ligand immobilization, leading blank injections, analyte injections, chip regeneration) which provide two duplicate measurement attempts per clone per run. In most experiments, technical replicates measured in separate runs further double the number of measurement attempts per clone to four.

Low Diversity Library Sequencing

To identify the DNA sequence of individual antibody variants evaluated by SPR, duplicate plates are provided for sequencing. A portion of the pelleted material is transferred into 96 well PCR (Thermo-Fisher) plate via pinner (Fisher Scientific) which contains reagents for performing an initial phase PCR of a two-phase PCR for addition of Illumina adapters and sequencing. Reaction volumes used are 12.5 μl. During the initial PCR phase, partial Illumina adapters are added to the amplicon via 4 PCR cycles. The second phase PCR adds the remaining portion of the Illumina sequencing adapter and the Illumina i5 and i7 sample indices. The initial PCR reaction uses 0.45 μM UMI primer concentration, 6.25 μl Q5 2x master mix (New England Biolabs) and PCR grade H_2O. Reactions are initially denatured at 98°C for 3 min, followed by 4 cycles of 98°C for 10 s; 59°C for 30 s; 72°C for 30 s; with a final extension of 72°C for 2 min. Following the initial PCR, 0.5 μM of the secondary sample index primers are added to each reaction tube. Reactions are then denatured at 98°C for 3 min, followed by 29 cycles of 98°C for 10 s; 62°C for 30 s; 72°C for 15 s; with a final extension of 72°C for 2 min. Reactions are then pooled into a 1.5 mL tube (Eppendorf). Pooled samples are size selected with a 1x AMPure XP (Beckman Coulter) bead procedure. Resulting DNA samples are quantified by Qubit fluorometer. Pool size is verified via Tapestation 1000 HS and is sequenced on an Illumina MiSeq Micro (2x150 nt) for HCDR3 libraries or an Illumina MiSeq Reagent Kit v3 (2x300 nt) for HCDR1-HCDR3 libraries with 20 % PhiX.

After sequencing, amplicon reads are merged using Fastp [90], trimmed by cutadapt [87] and each unique sequence enumerated. Next, custom R scripts are applied to calculate sequence frequency ratios between the most abundant and second-most abundant sequence in each sample. Levenshtein distance is also calculated between the two sequences. These values are used for downstream filtering to ensure a clonal population is measured by SPR. The most abundant sequence within each sample is compared to the designed sequences and discarded if it does not match any expected sequence. Dominant sequences are then combined with their companion Carterra SPR measurements.

Antibody Databases

The Observed Antibody Space (OAS) [11] was retrieved on February 1st, 2022. The Structural Antibody Database (SAbDab) [22] was retrieved on August 29th, 2022. The Coronavirus Antibody Database (CoV-AbDab) [80] was retrieved on December 21st, 2022.

To compute edit distance (number of mutations) between antibody sequences we use the Levenshtein distance, denoted lev. The "Minimum HCDR3 edit distance to OAS" is computed by taking the minimum edit distance between an HCDR3 of interest and all HCDR3s in OAS:

$$\min_{y \in \text{OAS}} \text{lev}(x_{\text{HCDR3}}, y_{\text{HCDR3}}).$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

This value is computed analogously for HCDR1 and HCDR2 as well as for other databases such as SABdAB, CoV-AbDab, or our training data. The "Minimum HCDR123 edit distance to OAS" is computed by taking the minimum edit distance between the tuple of HCDRs (HCDR1, HCDR2, HCDR3) belonging to an antibody of interest and all such tuples in OAS (computed analogously for other databases):

$$\min_{y \in \text{OAS}} \left(\text{lev}(x_{\text{HCDR1}}, y_{\text{HCDR1}}) + \text{lev}(x_{\text{HCDR2}}, y_{\text{HCDR2}}) + \text{lev}(x_{\text{HCDR3}}, y_{\text{HCDR3}}) \right).$$

To compute the OAS HCDR3 length distribution (Table S4) we iterate through all heavy chain sequences in OAS, consider the HCDR3 length, and maintain a tally of HCDR3 lengths. We then restrict to HCDR3 sequences with lengths between 9 and 17 and normalize to get the length distribution. For the OAS-J HCDR3 length distribution we do an analogous process but iterate only through heavy chain sequences in OAS that have trastuzumab's J-gene. For the SAbDab HCDR3 length distribution we take all unique HCDR3s in SAbDab with lengths between 9 and 17 and compute the frequencies at each length.

In Silico Structural Modeling

Three-dimensional models of selected *de novo* HCDR3 binders are created in PyMOL [91] and the Crystallographic Object-Oriented Toolkit (Coot) [92] using the coordinates of the trastuzumab-HER2 complex (PDB:1N8Z). Rosetta's FastRelax application [75] is applied using flexible backbone and side-chain degrees of freedom parameters. Prior to the relax procedure, we first idealize all candidate structures using Rosetta's Idealize protocol to avoid steric clashes and improper geometry. We relax using the maximum number of rotamers by passing *-EX1*, *-EX2*, *-EX3* and *-EX4* flags at initialization. We also include flags *-packing:repack_only* to disable design, *-no_his_his_pairE*, and *-multi_cool_annealer 10* to set the number of annealing iterations. For ranking of conformations in FastRelax, we use Rosetta's REF2015 energy function. It is well known that running relax on a structure will often move the backbone a few Angstroms[3], so we include an additional term containing harmonic distance constraints for all pairs of $C\beta$ atoms that are either not part of a CDR loop or not within distance 10 to any atom in a CDR loop, based on the conformation of the initial structure. These constraints are given weight 10^{-4}. The protocol is run ten times for each target, and we select the decoy with the lowest energy in the HCDR3 loop.

Multi-Step CDR Design

In an earlier version of this manuscript, we reported successful design of multiple CDRs using a multi-step generative AI design method, which we validated by SPR (Materials and Methods). We have deprecated this method internally in favor of our latest de novo methods. We report multiple binding designs to HER2 identified in a library of fewer than 500 multi-step designed multi-HCDR variants (Table S8). We find that these binders again are distinct from examples in the model's training data and antibodies in the SAbDab and OAS databases (Figure S13).

[3]Motivation is given in the official Rosetta documentation for Fast Relax

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

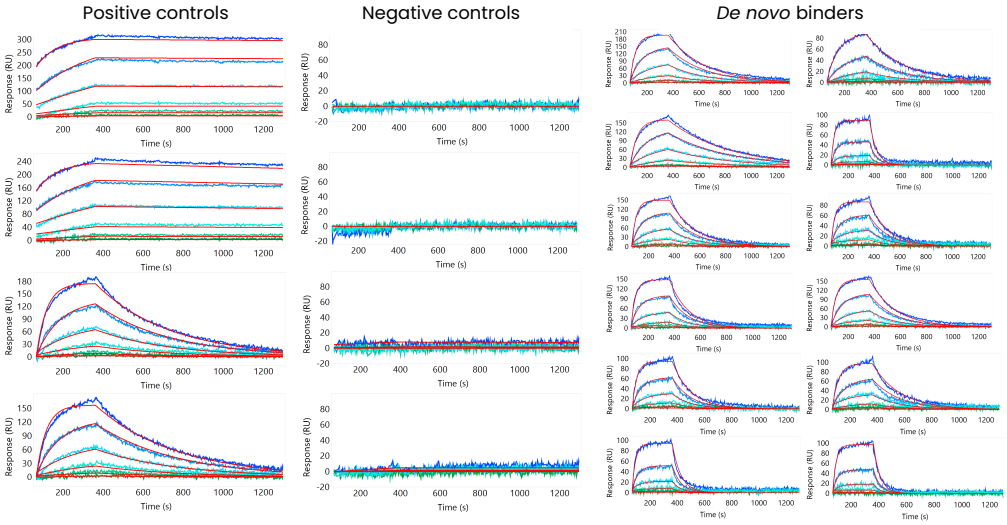

Fig S1. Sensorgram examples of high-throughput SPR workflow for identifying *de novo* binders. Two positive controls are shown, each with two replicates (top-most two show a high-affinity binder, bottom-most two show a low-affinity binder). Two negative controls are shown, each with two replicates. Two replicates are shown for each of six *de novo* binders (each row represents one binder).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

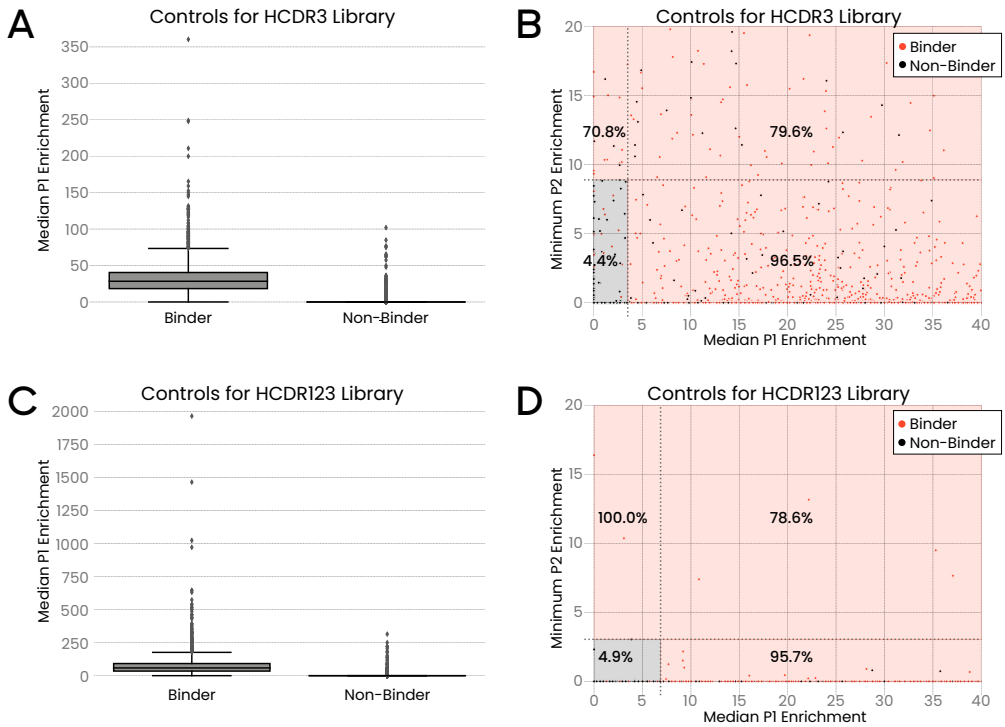

Fig S2. ACE enrichment scores are predictive of binding. Binders are classified in ACE based on median P1 enrichment and minimum P2 enrichment across three replicates (Materials and Methods). **(A)** Distribution of median P1 enrichment scores for HCDR3 controls separated by binders and non-binders. Binders have statistically significantly higher average median P1 enrichment than non-binders (Student's t-test, $t = 63.86, p < 10^{-10}$). **(B)** Plot of HCDR3 controls showing median P1 enrichment scores and minimum P2 enrichment scores. Sequences in the bottom left quadrant (shaded black) are labeled as non-binders whereas sequences in any of the other three quadrants (shaded orange) are labeled as binders. The percentage of sequences in each quadrant that are true binders (according to SPR) is shown. Axes truncated to enable better visualization. **(C)** Distribution of median P1 enrichment scores for HCDR123 controls separated by binders and non-binders. Binders have statistically significantly higher average median P1 enrichment than non-binders (Student's t-test, $t = 46.04, p < 10^{-10}$). **(D)** Plot of HCDR123 controls showing median P1 enrichment scores and minimum P2 enrichment scores. Sequences in the bottom left quadrant (shaded black) are labeled as non-binders whereas sequences in any of the other three quadrants (shaded orange) are labeled as binders. The percentage of sequences in each quadrant that are true binders (according to SPR) is shown. Axes truncated to enable better visualization.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

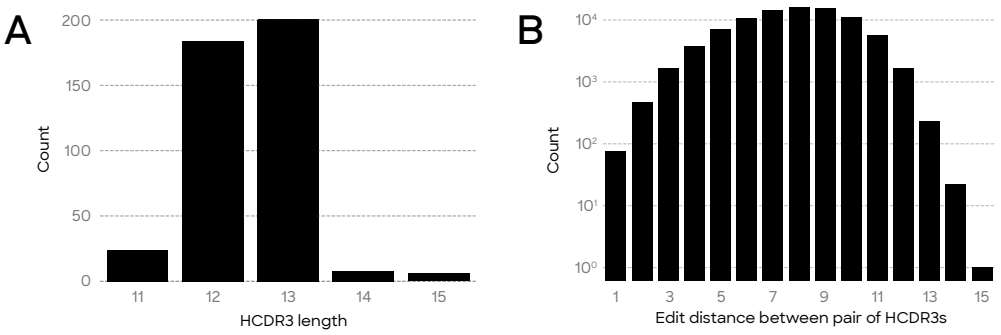

Fig S3. **(A)** Distribution of HCDR3 lengths for zero-shot-designed binders to HER2 (minimum of 11, maximum of 15, median of 13, mean of 12.5±0.69 SD). **(B)** Distribution (on log scale) of pairwise HCDR3 edit distances for zero-shot-designed binders to HER2.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

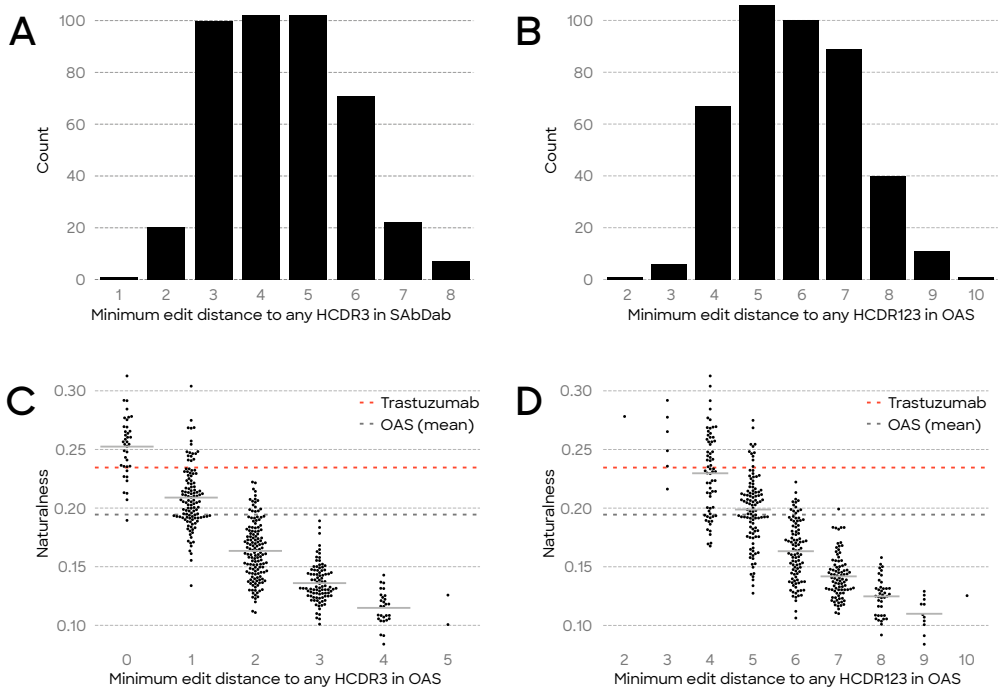

Fig S4. **(A)** Distribution of minimum edit distance to HCDR3s in SAbDab for zero-shot-designed binders to HER2 (minimum of 1, maximum of 8, median of 4, mean of 4.46 ± 1.37 SD) **(B)** Distribution of minimum edit distance to HCDR123s in OAS for zero-shot-designed binders to HER2 (minimum of 2, maximum of 10, median of 6, mean of 5.87 ± 1.38 SD) **(C)** Naturalness scores of zero-shot-designed binders vs. minimum edit distance to OAS HCDR3s. Naturalness score tends to decrease as distance to OAS increases, which is expected since the Naturalness model is trained on OAS sequences. Note that at minimum HCDR3 distance 0 to OAS the Naturalness scores of the designed binders are on average higher than trastuzumab's ($p < 2 \cdot 10^{-4}$) and for designed binders with minimum distance one or fewer HCDR3 mutations from OAS have higher average Naturalness scores than the OAS baseline ($p < 10^{-13}$). **(D)** Naturalness scores of zero-shot-designed binders vs. minimum edit distance to OAS HCDR123s. Note the presence of designs up to 5 HCDR123 mutations away from OAS with higher naturalness than trastuzumab. At 4 or fewer HCDR123 mutations to OAS the Naturalness scores of the designed binders are on average higher than the OAS basline ($p < 10^{-14}$).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

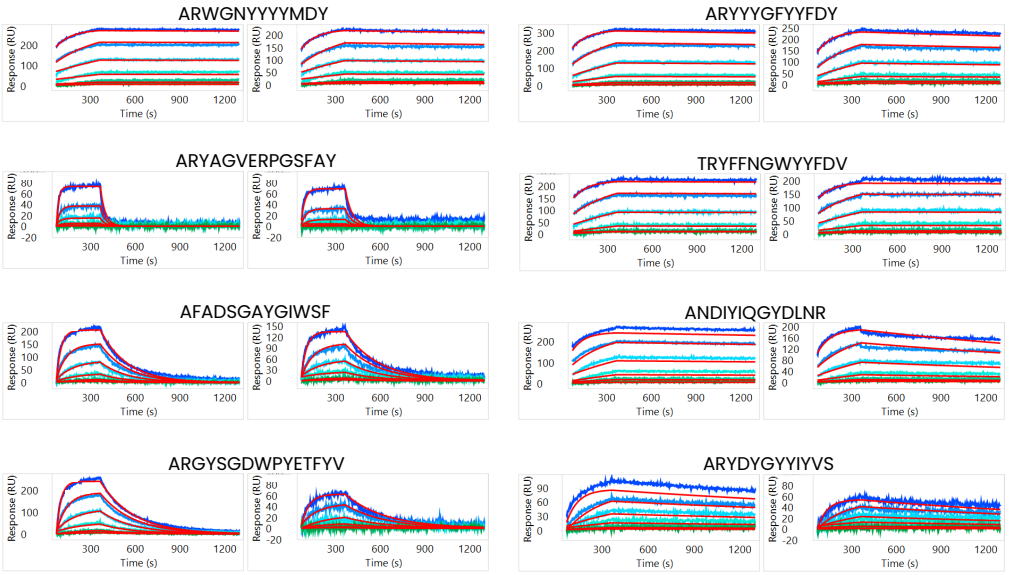

Fig S5. Sensorgrams of eight selected *de novo* HER2 binders. Each sensorgram represents two replicates of a single experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

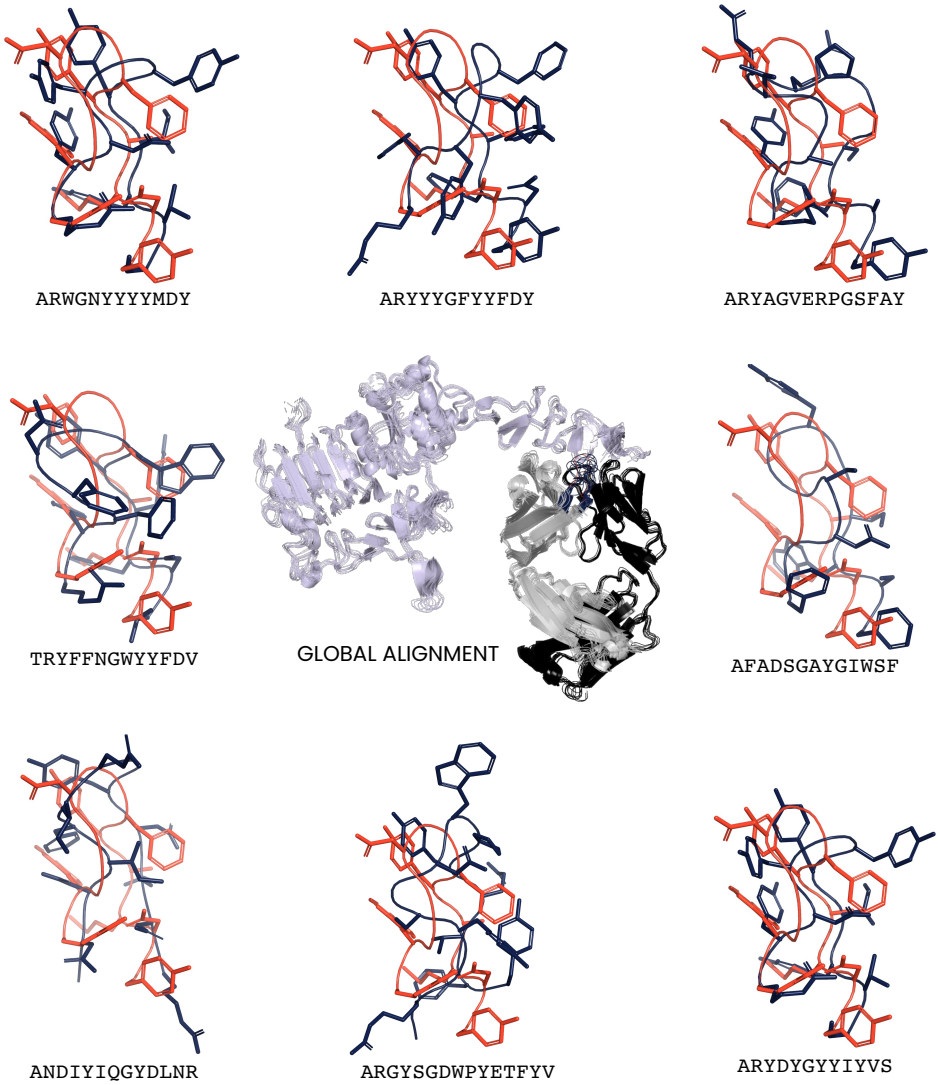

Fig S6. Conformational flexibility of *de novo* designed HCDR3s. Alignment of eight selected *de novo* HER2 binders with trastuzumab-HER2 complex shows small overall differences in the antigen (lavender), the heavy chain (gray) and the light chain (dark gray) structu re but large conformational changes in the HCDR3 regions. The trastuzumab HCDR3 loop is colored red and the *de novo* HCDR3 loops are colored blue.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

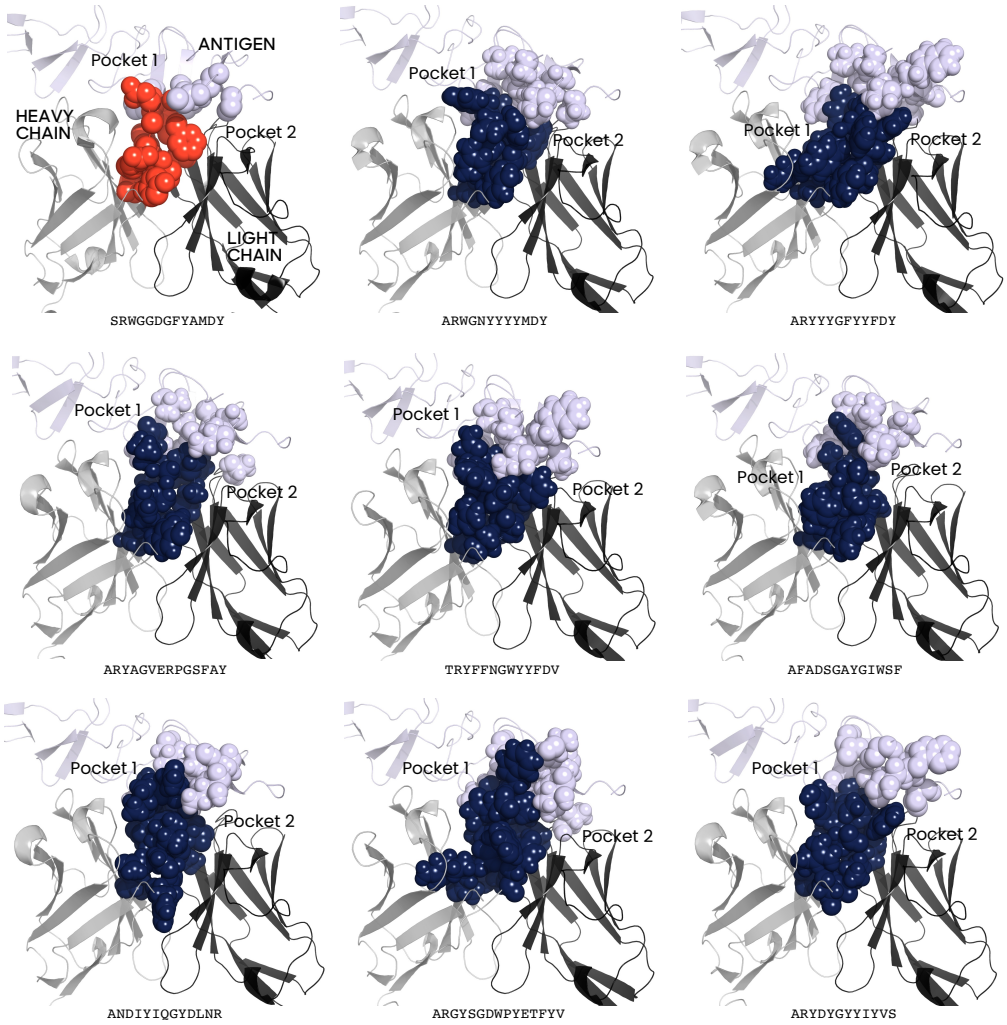

Fig S7. Space-filling representation of HCDR3 loops interacting with epitope residues. Residues were selected using a 5Å cutoff between HCDR3 and epitope residues. The trastuzumab HCDR3 loop is colored red (top left) and the *de novo* HCDR3 loops are colored blue. Two distinct epitope pockets that differentially interact with residues of each HCDR3 can be seen. HCDR3-epitope interacting surfaces vary based on HCDR3 sequence and conformation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

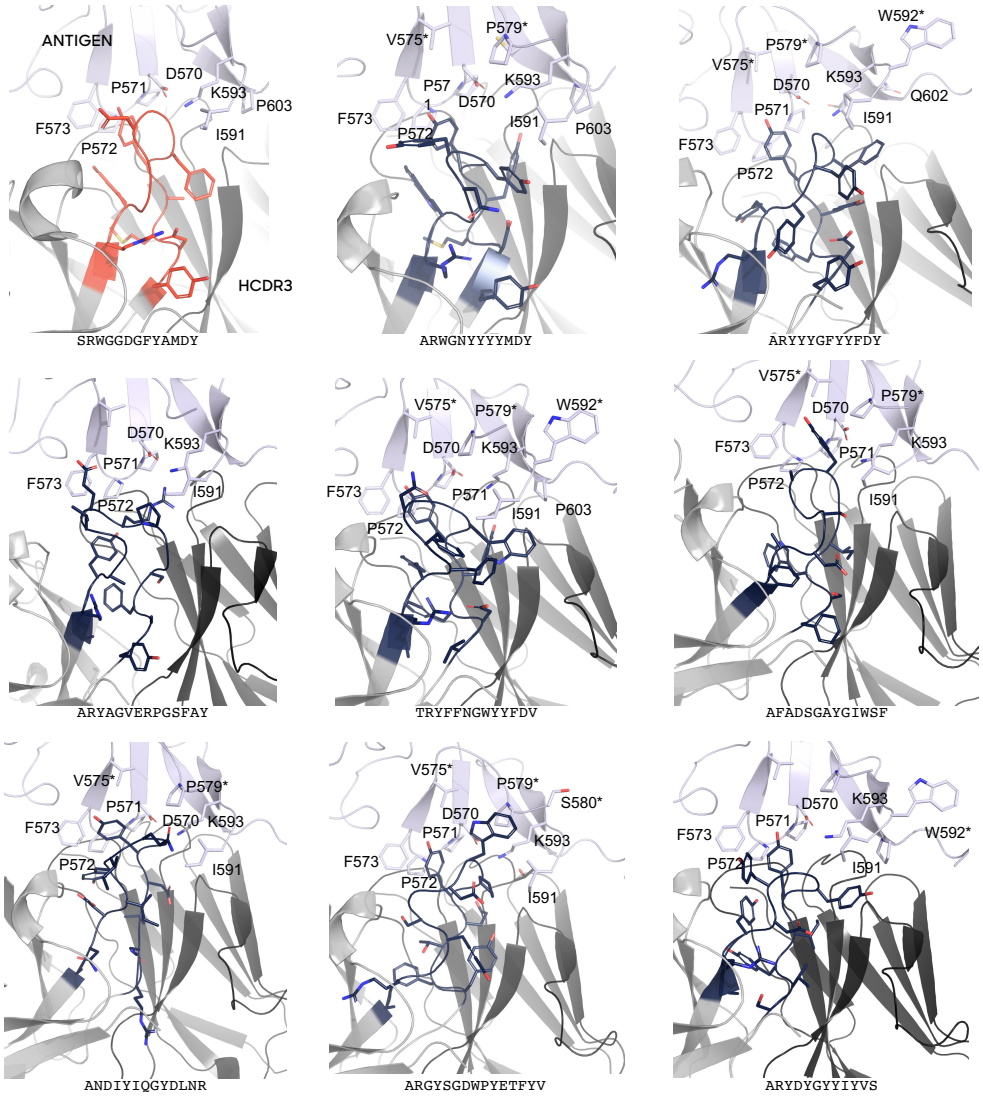

Fig S8. Stick representations of HCDR3-epitope interfaces. Residues were selected using a 5Å cutoff between HCDR3 and epitope residues (computed over all atoms). The trastuzumab HCDR3 loop is colored red (top left) and the *de novo* HCDR3 loops are colored blue. Epitope residues are labeled according to crystal structure PDB:1N8Z. An * denotes novel epitope residues in the *de novo* HCDR3 complexes that are not observed in the trastuzumab-HER2 complex.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ranibizumab	**A**	**K**	**Y**	**P**	**Y**	**Y**	**Y**	**G**	**T**	**S**	**H**	**W**	**Y**	**F**	**D**	**V**
De novo binder	**A**	**S**	**G**	**N**	**F**	**S**	**Y**	**P**	**D**	**D**	**Y**	**G**		**M**	**D**	**W**

Fig S9. Alignment of *de novo* designed VEGF binder to the HCDR3 of ranibizumab. The design is an HCDR3 variant of the therapeutic ranibizumab, which binds to human vascular endothelial growth factor A (VEGF-A) [78]. The designed binder has an affinity of 48.2 nM to VEGF-A, as measured by SPR, compared to sub-nanomolar binding of ranibizumab (0.37 nM). The designed HCDR3 is diverse and novel as it is 13 mutations away from ranibizumab's HCDR3 and at least 4 mutations away from any HCDR3 in OAS.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

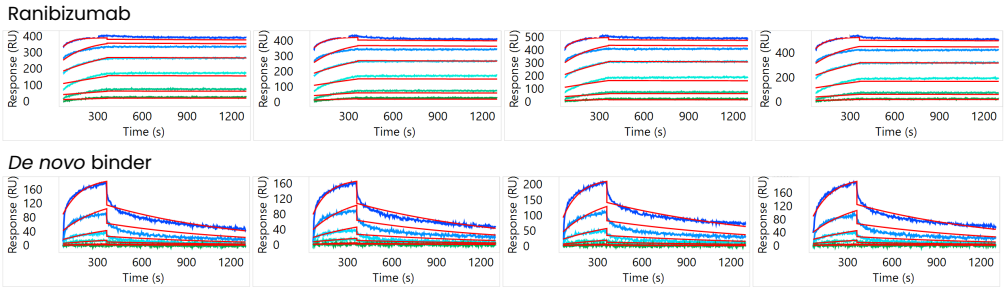

Fig S10. Sensorgrams of Ranibizumab Fab (positive control) and *de novo* designed HCDR3 binding to VEGF-A. Each sensorgram represents four replicates of a single experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

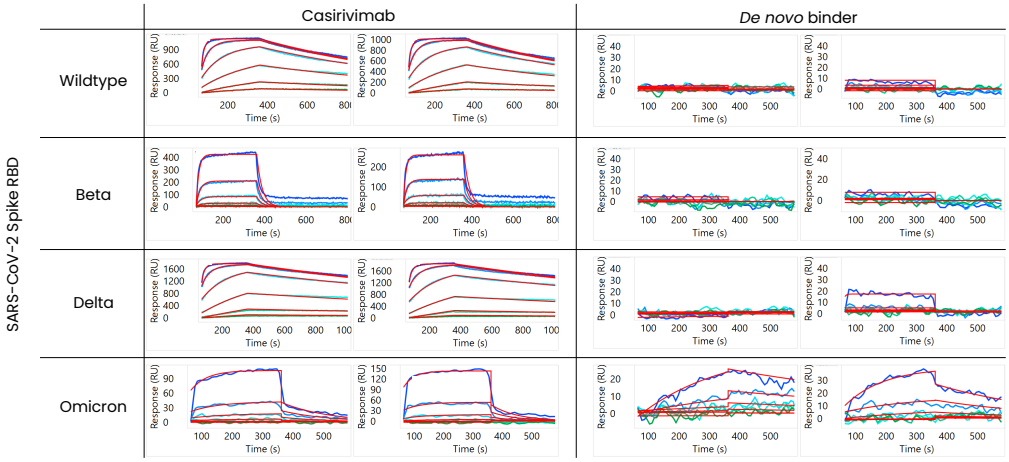

Fig S11. Sensorgrams of Casirivimab Fab (positive control) and *de novo* designed HCDR3 binding to SARS-CoV2 spike RBD variants. Each sensorgram represents two replicates of a single experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Casirivimab	**A**	**R**	**D**	**R**	**G**	**T**	**T**	**M**	**V**	**P**		**F**	**D**	**Y**
De novo binder	**A**	**R**	**D**	**W**	**G**	**Q**	**T**	**W**	**D**	**S**	**Y**	**F**	**D**	**Y**

Fig S12. Alignment of *de novo* designed SARS-CoV-2 Omicron binder to the HCDR3 of casirivimab. The designed HCDR3 is diverse and novel as it 6 mutations away from casirivimab's HCDR3, at least 2 mutations away from any HCDR3 in OAS, and at least 4 mutations away from any HCDR3 in CoV-AbDab (a database of antibodies capable of binding coronaviruses [80]).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

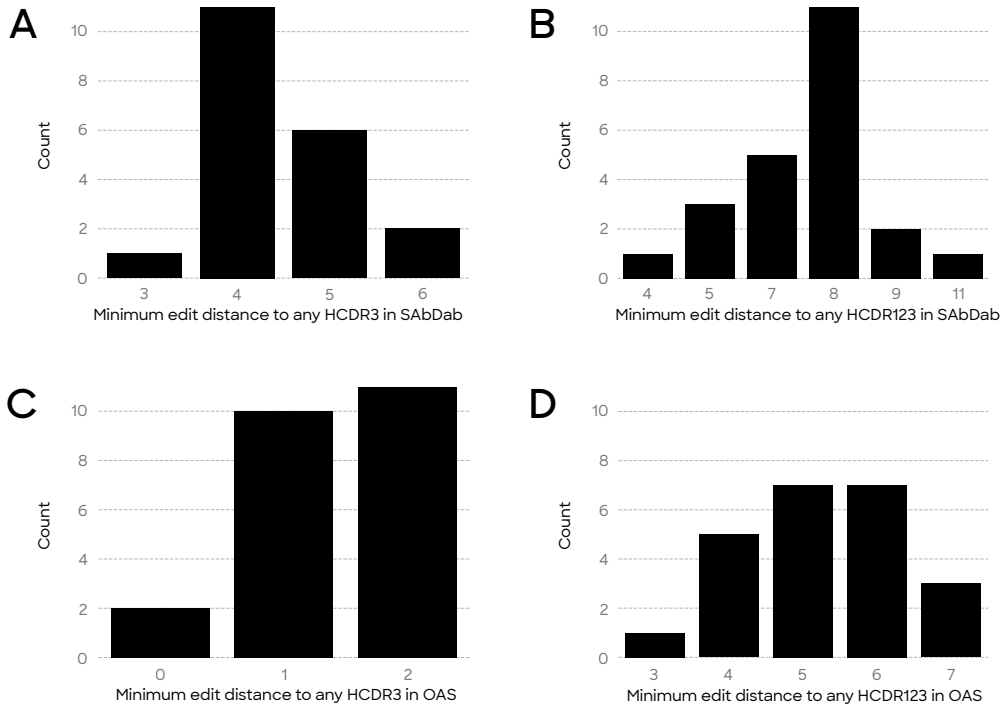

Fig S13. Multi-step multi-CDR design sequence analysis. (A) Distribution of minimum edit distance to HCDR3s in SAbDab for multi-step multi-HCDR AI-designed binders to HER2 (minimum of 3, maximum of 6, median of 4, mean of 4.39 ± 0.71 SD). Of the five designs with distinct HCDR1s from trastuzumab, three are at least one mutation away from SAbDab and two are contained in SAbDab. Similarly, of the 19 designs with distinct HCDR2s from trastuzumab, 15 are at least one mutation away from SAbDab and four are contained in SAbDab. This is expected given the fact that HCDR1 and HCDR2 display lower sequence diversity than HCDR3. **(B)** Distribution of minimum edit distance to HCDR123s in SAbDab for multi-step multi-HCDR AI-designed binders to HER2 (minimum of 4, maximum of 11, median of 8, mean of 7.43 ± 1.50 SD). Note the increase in mutations compared to HCDR3 edit distance despite the proximity of the HCDR1 and HCDR2 designs to SAbDab. **(C)** Distribution of minimum edit distance to HCDR3s in OAS for multi-step multi-HCDR AI-designed binders to HER2 (minimum of 0, maximum of 2, median of 1, mean of 1.39 ± 0.64 SD). All HCDR1 and HCDR2 designs are contained in OAS, which is again expected given their lower diversity. **(D)** Distribution of minimum edit distance to HCDR123s in OAS for multi-step multi-HCDR AI-designed binders to HER2 (minimum of 3, maximum of 7, median of 5, mean of 5.26 ± 1.07 SD). Note the increase in mutations compared to HCDR3 edit distance despite the HCDR1 and HCDR2 designs presence in OAS.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1. ACE Performance on HCDR3 Controls. Confusion matrix for controls indicating binding as measured by SPR and binding as measured by ACE. Accuracy, precision, recall, and F1-score shown below.

		ACE		
		Binder	**Non-Binder**	**Total**
SPR	**Binder**	2089	114	2203
	Non-Binder	101	2491	2592
	Total	2190	2605	4795

Accuracy	**Precision**	**Recall**	**F1**
95.52%	95.39%	94.83%	0.9511

Table S2. ACE performance on HCDR123 controls. Confusion matrix for controls indicating binding as measured by SPR and binding as measured by ACE. Accuracy, precision, recall, and F1-score shown below.

		ACE		
		Binder	**Non-Binder**	**Total**
SPR	**Binder**	2109	161	2270
	Non-Binder	96	3102	3198
	Total	2205	3263	5468

Accuracy	**Precision**	**Recall**	**F1**
95.3%	95.65%	92.91%	0.9426

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3. *De novo* designs achieve high, calibrated binding rates which outperform biological baselines. The top 100, 1,000, and 10,000 sequences by model likelihood are sampled and tested experimentally. As the number of selected sequences goes up, binding rate goes down. This indicates that the model's likelihood is calibrated with binding. N/A indicates that fewer than 10,000 sequences were sampled for experimental testing. * $p < 0.01$, Fisher's exact tests with each of the biological baselines (OAS, OAS-J, and SAbDab). † $p < 0.01$, Fisher's exact tests with each of the "Wrong Antigen" populations (rat HER2, HER3, VEGF).

Method	**HCDR3 Binding Rate (%)**			**HCDR123 Binding Rate (%)**		
	$k = 100$	$k = 1000$	$k = 10000$	$k = 100$	$k = 1000$	$k = 10000$
De novo (Human HER2)	**13***	**10.6*†**	**6.99*†**	**3***	**1.8*†**	**1.33***
Wrong Antigen (Rat HER2)	0	2.8	1.49	0	0.5	N/A
Wrong Antigen (HER3)	4	2.9	1.27	0	0.2	N/A
Wrong Antigen (VEGF)	4	2.5	1.21	0	0	N/A

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4. OAS, OAS-J, and SAbDab HCDR3 length distributions. Frequency of heavy chain CDR3s appearing in OAS with lengths between 9 and 17 amino acid residues are shown. OAS-J frequency adds the condition that HCDR3s belong to an antibody with trastuzumab's J-gene. Frequency of unique heavy chain CDR3s in SAbDab with lengths between 9 and 17 residues.

HCDR3 Length	9	10	11	12	13	14	15	16	17
Frequency (OAS)	0.042	0.067	0.105	0.128	0.145	0.156	0.142	0.117	0.099
Frequency (OAS-J)	0.042	0.089	0.141	0.156	0.165	0.163	0.119	0.077	0.048
Frequency (SAbDab)	0.097	0.081	0.125	0.145	0.114	0.130	0.119	0.100	0.089

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5. Number of binders and number of sequences tested experimentally for biological baseline populations.

Population	# HCDR3 Binders	# HCDR3s	# HCDR123 Binders	# HCDR123s
OAS	1,334	49,766	79	49,346
OAS-J	521	9,921	32	9,862
SAbDab	75	2,374	1	1,553
Permuted Sequences	16	4,868	N/A	N/A

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S6. Mean Naturalness scores across different groups (using grafting into trastuzumab scaffold) and p-values when comparing to the *de novo* binders.

Group	**Mean Naturalness**	**p-value (comparison to *de novo*)**
***De novo* binders**	0.175	N/A
Trastuzumab	0.235	$1.20 \cdot 10^{-94}$
OAS	0.194	$3.33 \cdot 10^{-16}$
Frequency baseline	0.143	$1.19 \cdot 10^{-52}$
Phage display baseline	0.136	$4.21 \cdot 10^{-87}$
Scrambled OAS	0.101	$1.33 \cdot 10^{-298}$

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S7. Measured binding affinities of casirivimab and *de novo* binder variant to SARS-CoV-2 antigens. N/A indicates a lack of binding. In this setting, the model was instructed to generate HCDR3 variants targeting the Omicron variant of COVID that fit within the casirivimab framework. These results indicate that the model successfully designed an HCDR3 variant that loses activity to SARS-CoV-2 Spike RBD Wildtype/Beta/Delta, yet maintains activity to Omicron. These results are a first step toward controllability of antibody design to specific protein variants.

	Wildtype	**Beta**	**Delta**	**Omicron**
Casirivimab	8.3nM	470.0nM	4.2nM	240.0nM
***De novo* binder**	N/A	N/A	N/A	179.7nM

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.08.523187; this version posted March 29, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S8. Multi-step AI-designed trastuzumab variant binders to HER2 with all three HCDRs designed. ED indicates edit distance from an HCDR to the corresponding trastuzumab HCDR. Note the model occasionally recovers the native trastuzumab HCDR1 and HCDR2. We display nine variants here and open source the entire set of 23 designs in accompanying sequence data. We have deprecated this multi-step design protocol internally, in favor of our fully *de novo* approach.

	HCDR1 (ED)	HCDR2 (ED)	HCDR3 (ED)	KD (nM)
Trastuzumab	GFNIKDTY (0)	IYPTNGYT (0)	SRWGGDGFYAMDY (0)	1.94
AI Design 1	GFNVSSSY (4)	IYPENGST (2)	SRYSYYGGYAFDY (6)	224.9
AI Design 2	GFNISDYY (2)	IYPSSGYT (2)	ARDGGVGSYSMDV (6)	1323.0
AI Design 3	GFNIKDYY (1)	IYPSRGYT (2)	ARHGSEGRYTMDV (7)	192.1
AI Design 4	GFNIKDTW (1)	IDPANGYT (2)	ARVGGRGYYYYDY (6)	394.1
AI Design 5	GFNFKDTY (1)	IYPANGYT (1)	ASDGGSYSYAFDY (7)	642.2
AI Design 6	GFNIKDTY (0)	IDPADGYT (3)	AREGGYGYYAMDY (4)	160.1
AI Design 7	GFNIKDTY (0)	IDPANGYT (2)	ARVGGRGYYYYDY (6)	134.0
AI Design 8	GFNIKDTY (0)	IYPSNGYT (1)	ARYGGSYAYYFYY (8)	18.6
AI Design 9	GFNIKDTY (0)	IYPTNGYT (0)	ARYYGSSSYYFAY (9)	60.6