

1 Lab-in-the-loop therapeutic antibody design with
2 deep learning

3 Nathan C. Frey^{1*†}, Isidro Hötzl^{1†}, Samuel D. Stanton^{1†},
4 Ryan Kelly^{1†}, Robert G. Alberstein^{1†}, Emily Makowski¹,
5 Karolis Martinkus¹, Daniel Berenberg¹, Jack Bevers III¹,
6 Tyler Bryson¹, Pamela Chan¹, Alicja Czubaty²,
7 Tamica D'Souza¹, Henri Dwyer¹, Anna Dziewulska¹,
8 James W. Fairman¹, Allen Goodman¹, Jennifer Hofmann¹,
9 Henry Isaacson¹, Aya Ismail¹, Samantha James¹, Taylor Joren¹,
10 Simon Kelow¹, James R. Kiefer¹, Matthieu Kirchmeyer¹,
11 Joseph Kleinhenz¹, James T. Koerber¹, Julien Lafrance-Vanasse¹,
12 Andrew Leaver-Fay¹, Jae Hyeon Lee¹, Edith Lee¹, Donald Lee¹,
13 Wei-Ching Liang¹, Joshua Yao-Yu Lin¹, Sidney Lisanza¹,
14 Andreas Loukas¹, Jan Ludwiczak¹, Sai Pooja Mahajan¹,
15 Omar Mahmood¹, Homa Mohammadi-Peyhani¹, Sanrupti Nerli¹,
16 Ji Won Park¹, Jaewoo Park¹, Stephen Ra¹, Sarah Robinson¹,
17 Saeed Saremi¹, Franziska Seeger¹, Imee Sinha¹, Anna M. Sokol²,
18 Natasa Tagasovska¹, Hao To¹, Edward Wagstaff¹, Amy Wang¹,
19 Andrew M. Watkins¹, Blair Wilson¹, Shuang Wu¹,
20 Karina Zadorozhny¹, John Marioni¹, Aviv Regev¹, Yan Wu¹,
21 Kyunghyun Cho¹, Richard Bonneau^{1*}, Vladimir Gligorijević^{1*}

22 ¹Genentech.

23 ²Roche.

24 *Corresponding author(s). E-mail(s): frey.nathan.nf1@gene.com;
25 bonneau.richard@gene.com; gligorijevic.vladimir@gene.com;

26 †These authors contributed equally to this work.

27 Abstract

28 Therapeutic antibody design is a complex multi-property optimization problem
29 that traditionally relies on expensive search through sequence space. Here, we
30 introduce “Lab-in-the-loop,” a new approach to antibody design that orchestrates
31 generative machine learning models, multi-task property predictors, active learn-
32 ing ranking and selection, and *in vitro* experimentation in a semi-autonomous,
33 iterative optimization loop. By automating the design of antibody variants, prop-
34 erty prediction, ranking and selection of designs to assay in the lab, and ingestion
35 of *in vitro* data, we enable a holistic, end-to-end approach to antibody optimiza-
36 tion. We apply lab-in-the-loop to four clinically relevant antigen targets: EGFR,
37 IL-6, HER2, and OSM. Over 1,800 unique antibody variants are designed and
38 tested, derived from lead molecule candidates obtained via animal immunization
39 and state-of-the-art immune repertoire mining techniques. Four lead candidate
40 and four design crystal structures are solved to reveal mechanistic insights into
41 the effects of mutations. We perform four rounds of iterative optimization and
42 report 3–100× better binding variants for every target and ten candidate lead
43 molecules, with the best binders in a therapeutically relevant 100 pM range.

44 **Keywords:** Antibody design, deep learning, lab-in-the-loop, drug discovery

45 1 Main

46 Discovery of therapeutics is a fundamental driver of the extension of human life and
47 health-span. Antibodies represent one of the most versatile classes of therapeutic
48 modalities. The genetic modularity and biological mechanisms for rapid diversification
49 underlying their production and maturation enable the jawed vertebrate immunolog-
50 ical response to recognize almost any target molecule of interest, including proteins,
51 peptides, and small molecules [1, 2]. As therapeutics, antibodies can be leveraged for
52 immune system engagement or to deliver drug payloads with high specificity, creating
53 strong demand for new antibodies to attack disease mechanisms at the source, leading
54 to 58 new FDA-approved antibody therapeutics from 2019-2023 [3]. Unfortunately,
55 antibody discovery and engineering remains a laborious and challenging process,
56 adding to high development costs that influence the final cost of these medicines.

57 Antibody discovery campaigns commonly begin with immunization of animals
58 with the target antigen, eliciting an *in vivo* immune response that can be quanti-
59 fied by analysis of the B-cell repertoire before and after immunization, typically by
60 hybridoma or single B cell screening [4]. Alternatively, large libraries of 10^9 - 10^{11} naive
61 antibody sequences can be expressed via display technologies (e.g., yeast or phage)
62 and panned against the antigen directly *in vitro* [5]. Molecules must be optimized
63 for antigen binding affinity, functional activity, immunogenicity, pharmacokinetics,
64 expression, stability, solubility, viscosity, and aggregation propensity [6, 7]. The high-
65 dimensional search space for antibody sequences has largely precluded first-principles
66 approaches to antibody design, and the large cost, long timelines, and linearity of
67 existing processes are prohibitive for the rapid development of new therapeutics [8]
68 and cannot effectively manage the balance of multiple properties.

Machine learning (ML) and Bayesian Optimization (BayesOpt) are natural approaches for finding therapeutic antibodies in the high-dimensional search space of sparsely functional proteins [9–15]. Language models trained on protein sequence data have been used to guide directed evolution to affinity mature antibodies [16–19]. However, for drug discovery campaigns, ML methods must respect unique constraints for therapeutic molecules (e.g., developability, specificity, potency, expression, efficacy) and augment or outperform state-of-the-art molecular engineering approaches [6]. Prior work in ML for antibody design shows the immense promise of these approaches [16, 20]. However, in our hands these methods result in antibody variants that do not affect binding affinity and do not diversify the lead candidates, or introduce germline-reverting mutations that abrogate affinity. For therapeutic antibodies, we may require $\geq 3\times$ affinity maturation of an already matured antibody, with carefully chosen mutation locations to minimize immunogenicity risk. Unlike prior work, our optimization campaigns begin with reasonably optimized and viable clones from state-of-the-art *in vivo* discovery campaigns. Because of these stringent requirements and the complexity of the multi-property optimization problem outlined above, *de novo* protein discovery methods [21–23] do not obviate the need for optimization. Our approach takes steps towards respecting important therapeutic constraints [6], demonstrates generalizability across antigen targets and epitopes, and enables autonomous antibody engineering. Here, to address the limitations of both traditional discovery/engineering and prior ML-based design of antibodies, we introduce the “Lab-in-the-loop” (LitL) system.

We report results for four therapeutically relevant and biologically interesting antigen targets: Epidermal growth factor receptor (EGFR) [24], Interleukin-6 (IL-6) [25], Human epidermal growth factor receptor 2 (HER2) [26], and Oncostatin M (OSM) [27]. We discover at least $3\times$ better binding molecules to every single target, starting with lead antibodies from animal immunization hit discovery campaigns, and 10–100 \times better binders for multiple molecules. By leveraging an ensemble of generative models coordinated by a property prediction model for ranking and selection of designs, we achieve improvements across affinity, expression, and developability. We rationalize our results via experimental structure determination of 8 lead candidates and designs, as well as biophysics-based computational modeling of binding design complexes, showing that our designs preserve important interactions responsible for binding, while introducing up to eight new mutations that improve binding affinity. Our results demonstrate the powerful generalization capabilities of LitL to perform antibody design across diverse antigen targets and epitopes, without human intervention, while producing real therapeutic antibodies that are viable candidates to progress in the drug discovery pipeline.

2 Results

2.1 Lab-in-the-loop is a general machine learning system for autonomous drug discovery

LitL is a new approach for drug design that orchestrates loosely coupled ensembles of generative models, property prediction oracles, a ranking and selection algorithm, and experimental assays, all encompassed in an active learning loop to achieve iterative

optimization of antibody lead candidates (Fig 1a). Unlike prior work on biomolecular design with ML, which does not reflect the actual constraints and rigorous standards for success needed in drug discovery campaigns [6, 28, 29], LitL performs multi-property optimization, starting from lead candidate molecules, and operates across the human proteome of druggable antigen targets. In this paper, we focus on therapeutic antibody sequence space, although LitL is extensible to other therapeutic modalities. The selected example targets of EGFR, IL6, HER2, and OSM are of great interest in the treatment of cancer, inflammation, and autoimmune diseases [24–27].

Table 1: Lab-in-the-loop statistics.

Design round	Total designs	Expressors (> 0.01 mg)	Binders (> $4pK_D$)	Better binders (> 3×)
1	464	343	109	1
2	520	499	422	23
3	605	527	415	118
4	267	259	231	60

2.1.1 Generative models produce libraries of candidate molecules

Starting from a seed molecule that we aim to improve, the loop begins with the creation of a diverse library of sequences (i.e., the variants of the seed sequence) that is rich in high-quality candidate molecules. Generative models are trained on protein sequence data with the aim of emitting novel, diverse, functional proteins that are improved relative to sequences from *in vitro* or *in vivo* discovery and design campaigns. While *in vitro* and *in vivo* methods are restricted to sequences that are present in the display library or animal immune repertoire (which may require significant engineering before they are suitable for clinical trials), ML methods can generate near-limitless sequence diversity, optimize sequences with respect to multiple desirable properties, and enforce constraints on chemical liabilities and other molecular assessment criteria that would otherwise prevent molecules from advancing to clinical trials.

Generative models have various trade offs, generally between sample quality and sample diversity. To address the limitations of any particular generative model, we developed purpose-built methods for unguided sampling, sequence diversification, and hit expansion (e.g., Walk-Jump Sampling [30]) and separate generative methods for guided multi-property optimization (e.g., LaMBO-2 [31], DyAb [32], and Property Enhancer [33]). We include design proposals from multiple methods in each round, with the purpose of diversifying our design pool and maximizing our chances of success (as each method explores different parts of the antibody sequence space; see Supplementary Figures A4, A5, A6, A7).

We adaptively allocate experimental capacity to each method by combining the output of each method into a single combined library that is subsequently filtered and ranked by our active learning framework (i.e., global selection). In each round, all generative methods may produce up to 30,000 designs per lead molecule. This constraint is imposed to prevent any one method from dominating global selection

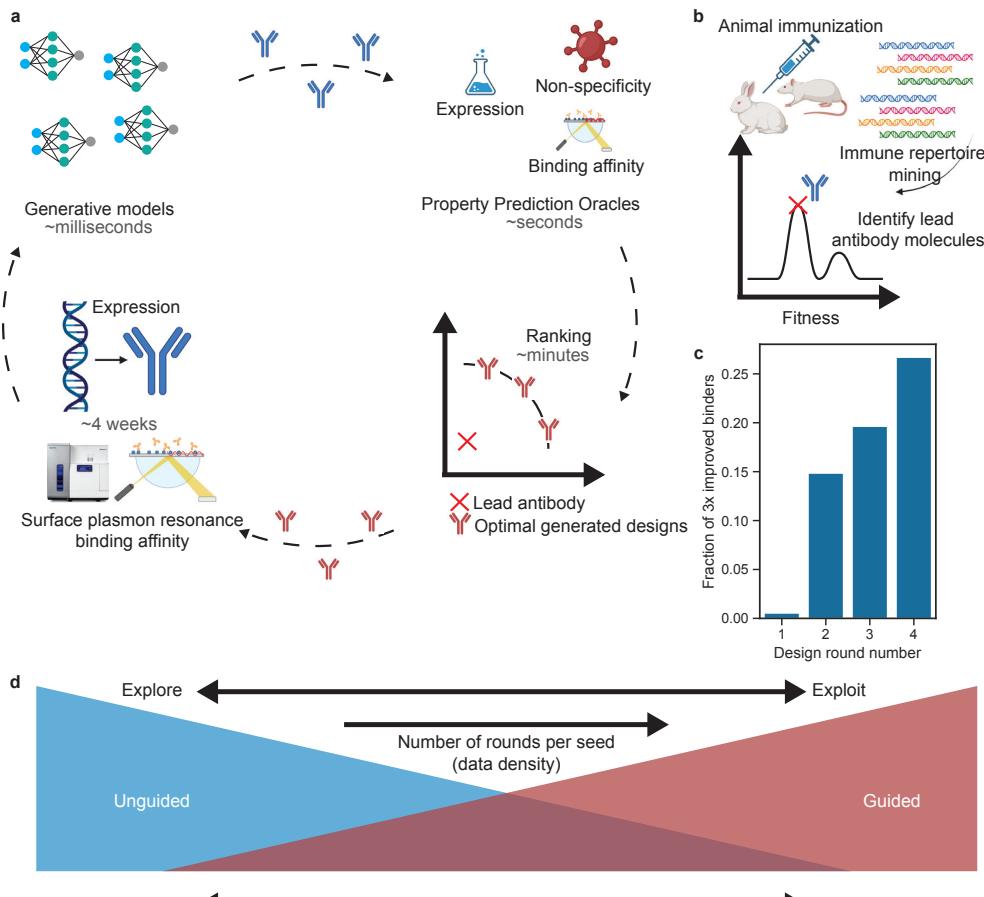


Fig. 1: Lab-in-the-loop autonomous therapeutic antibody optimization. **a**, Generative models produce a diverse *in silico* library of antibody sequence designs. Designs are ranked by property prediction oracles that are chosen, tournament style, based on their leaderboard performance evaluated on a hold-out split from the previous round of experiments. Predicted properties are used to rank designs to maximize the expected improvement compared to the lead antibody. Designs are expressed and binding affinity is measured to antigen targets with surface plasmon resonance (SPR). Lab data is used to re-train generative models and property prediction oracles for active learning. **b**, Lead antibody molecules are the best identified starting points from *in vivo* animal immunization and immune repertoire mining, although our approach can be initialized from any antibody discovery method. **c**, Lab-in-the-loop achieves monotonically increasing improvements in generating tighter binding variants over successive rounds of design. **d**, Ensembling unguided and guided generative models allows the LitL framework to dynamically explore and exploit regions of antibody sequence space.

146 by brute force, and to limit the computational cost of downstream *in silico* quality
147 control. Details of these generative methods, their architectures, and training datasets,
148 are presented in Section 4.1.1 of the Methods and their respective references.

149 2.1.2 Diverse molecule libraries are filtered and ranked for active 150 learning selection

151 Given a high-quality library of designs, we then accurately estimate their therapeutic
152 properties to determine which sequences to synthesize. In particular, generated designs
153 are annotated with predicted properties encompassing expression, binding affinity,
154 and non-specificity. All properties have associated binary labels (e.g., 1:1 binding to
155 target antigen is or is not detected by SPR at a fixed concentration) modeled by
156 binary classifiers, and expression yield and binding affinity are modeled as scalars by
157 regression models. Out-of-distribution (OOD) detection methods introduced in [30]
158 and additional quality control filters (e.g., chemical liability motifs) are also used to
159 assist the discriminator by excluding very anomalous sequences or those with easily
160 identified developability issues. Further details of these oracles, OOD detection meth-
161 ods, and quality control filters are presented in Sections 4.1.3, 4.1.4, and 4.1.5 of the
162 Methods and their respective references.

163 The predicted properties for each design that passes quality control filters are
164 then used to produce a global ranking of the designs that are predicted to perform
165 best in laboratory experiments. The ranking model is the global arbiter of design
166 selection. This systems-level design allows LitL to incorporate many diverse ideas and
167 methods, while ensuring that all experimental successes contribute to improving the
168 ranking model and oracles, which in turn improves the generative models, creating
169 a flywheel of continuous improvement. The ranking algorithm picks designs that are
170 *non-dominated*, that is, designs on the Pareto frontier of property space. We want
171 to avoid expensive experimental assays for designs that are strictly dominated, i.e.,
172 designs for which we already have labeled antibody candidates with properties at least
173 as good or better in all respects than the candidate design.

174 We select designs using the Noisy Expected Hypervolume Improvement (NEHVI)
175 *acquisition function* [34] to mitigate the effect of property measurement noise, auto-
176 matically induce diversity in sequence space that corresponds to diversity in property
177 space, and balance the explore-exploit tradeoff induced by the conflicting objectives
178 of improving the generalization capabilities of the discriminative models while still
179 producing high-quality candidate molecules as soon as possible. Further details on the
180 acquisition function and ranking are given in Section 4.1.6 of the Methods.

181 2.1.3 Experimental workflows drive the active learning loop

182 LitL is enabled by an efficient experimental pipeline capable of producing hundreds of
183 high-resolution data points in parallelized, overlapping 4-6 week timelines. To enable
184 the accelerated timelines of LitL, we developed an optimized antibody production
185 pipeline utilizing Gibson-assembled linear fragments (GLFs), as previously described
186 [35]. Briefly, variable region sequences are synthesized as linear DNA fragments (Twist
187 Bioscience) and subsequently integrated into GLFs containing all elements required

188 for transient expression at the 1mL scale utilizing HEK293 cells. The linear DNA
189 based expression workflow trims multiple cloning steps, saving significant time and
190 cost while also being highly amenable to automation. After expression, antibodies are
191 affinity purified and concentration is measured by optical density. Binding affinity of
192 the designed and lead antibodies is measured by surface plasmon resonance (SPR).
193 Further details on experimental methods are presented in Section 4.4 of the Methods.

194 **2.1.4 Lead antibody candidates are obtained from *in vivo*
195 discovery campaigns**

196 The lead antibody candidates are obtained from state-of-the-art *in vivo* immunization
197 of genetically engineered animals and immune repertoire mining techniques (Fig. 1b)
198 [4]. Repertoire mining produces natural, developable lead molecules and additionally
199 provides unlabeled local neighborhoods (in sequence space) of likely binders around
200 lead antibodies, which is invaluable for initial training sets of generative and property
201 prediction models. However, the LitL approach is fully general and starting seeds can
202 also be obtained from rationally designed yeast surface display [36] libraries and *in*
203 *vitro* discovery campaigns. Further details on seed selection are presented in Section
204 4.4.2 of the Methods.

205 **2.2 Lead candidates are affinity matured via iterative
206 optimization**

207 Before presenting the detailed sequence-level results of LitL, we highlight the top level
208 experimental success of the LitL approach. For simplicity, our top line metric of success
209 for a design is a strict threshold of at least 3× improvement in measured binding
210 affinity compared to the lead antibody. This unambiguously establishes a criterion for
211 “better binding”, beyond any uncertainty induced by measurement noise. In Fig. 1c,
212 we show the fraction of total designs that are 3× better binders discovered in each
213 round. In the first round of LitL *no labeled data is available* in the neighborhoods of
214 the starting seeds. By the fourth round, LitL produces a design library of >26% 3×
215 better binders. Figure A1 shows the cumulative number of labeled designs for each
216 seed over design rounds.

While affinity maturation is not our only optimization goal, it is a major focus of therapeutic antibody design to take strong binding, developable lead antibodies and engineer at least nanomolar binding before promoting leads to later stage development. For convenience, we represent binding affinity as a negative log transform of the equilibrium dissociation constant,

$$pK_D = -\log_{10}(K_D) = -\log_{10}(k_d/k_a) \quad (1)$$

217 where k_d is the dissociation rate (off rate) and k_a is the association rate (on rate). Fig.
218 2a shows a histogram of all LitL designs and their difference in binding affinity (ΔpK_D)
219 compared to their respective lead antibodies. The dashed vertical line indicates a 3×
220 improvement in binding affinity, such that all designs to the right of the line are at least
221 3× better binders ($\Delta pK_D = 0.47712$) than their respective lead molecules. A similar

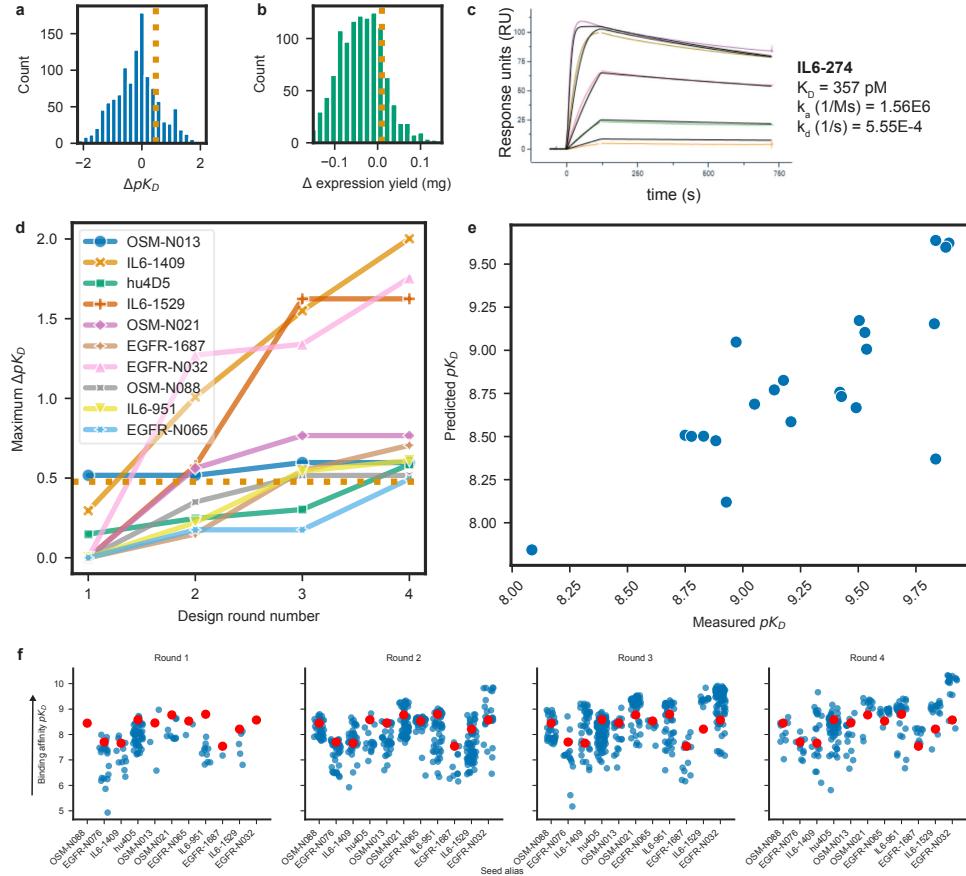


Fig. 2: Therapeutic antibody lead optimization via affinity maturation with lab-in-the-loop. **a**, Histogram of change in binding affinity (ΔpK_D) for all designed antibodies. The dashed vertical line denotes a $3\times$ improvement in binding affinity. **b**, Histogram of change in expression yield (mg) for all designed antibodies. The dashed vertical line denotes a 0.01 mg improvement. **c** Representative SPR sensogram for a 357 pM binder to IL6. **d**, Max improvement in binding affinity for 10 different lead antibodies across four different antigen targets. The dashed horizontal line indicates a $3\times$ improvement in binding affinity. **e**, Predicted versus measured pK_D for 22 of the highest binding affinity designs. **f** Absolute measured pK_D (higher is better) of designs (blue) and lead molecules (red) over design rounds.

histogram for expression yield is shown in 2b, where the dotted line corresponds to 0.01 mg, the amount of antibody required to run SPR. An example SPR sensogram binding curve (Figure 2c) illustrates how the affinity values are derived from curve fits. Across four antigen targets and ten distinct lead molecules¹, the improvement in pK_D for the

¹We considered 12 total starting seed molecules. One of these seeds failed to express. Designs from one seed underperformed the rest in the first round of iterative design. Our active learning framework (acquisition

best LitL design over rounds is shown in Fig. 2d. The horizontal dashed line indicates a $3\times$ improvement in binding. For ten lead molecules, LitL successfully discovers better binders that meet our pre-specified threshold for significant affinity maturation. For four of the lead molecules, LitL discovers better binders with $\Delta pK_D > 1$, which denotes a $10\times$ increase in binding affinity. In Figure 2e we show predicted versus measured pK_D for 22 of the tightest binding designs derived from each seed, across over a log-fold difference in binding affinities. Figure 2f shows the absolute measured pK_D for designs (in blue) over rounds and their respective seeds (in red).

2.3 Multi-property optimization enables therapeutic antibody design

In addition to initializing our optimization efforts from developable, natural antibodies derived from *in vivo* immune repertoire mining, we further constrain our designs to account for non-specificity, expression yield (which impacts downstream characterization assays that require specific amounts of protein material), and *in silico* developability risks. In Fig 3a we show the lead candidate molecules in dark blue and the best binding variant designs for each seed in light blue, after computing *in silico* metrics with the *Therapeutic Antibody Profiler* (TAP) [37]. These metrics, derived from the variable domain antibody sequence, flag sequences that are out-of-distribution with respect to the *Therapeutic Structural Antibody Database* (Thera-SAbDab). All but one of our best binding designs are within the TAP guideline ranges; one design is just over the threshold for positive and negative charge in the CDRs. LitL avoids drastic biochemical and biophysical changes that would induce avoidable developability risks, while achieving many-fold improvements to binding affinity.

To estimate risk of non-specific binding interactions of our designs, we developed a surrogate model trained on internal BV ELISA data. BV ELISA [38] is an assay developed to assess non-specific binding (further details are provided in Section 4.4). A BV ELISA score > 1.0 indicates a non-specificity risk; we confirm that all of our designs fall below the predicted risk threshold (Fig 3c). For 67 designs, we also engineer out chemical liabilities while maintaining or improving binding affinity (Fig A8). Chemical liabilities include deamidation ('NG') and glycosylation ('N*[ST]') motifs, and other easily recognizable substrings that pose developability risks. The per-round Pareto frontier of designs with respect to expression concentration and binding affinity (Fig 3d) illustrates how successive design rounds expand the Pareto frontier, finding non-dominated designs.

2.4 Modeled interface structures of designed antibodies demonstrate epistatic interactions

As all designs reported in this work were produced with sequence-only models, we considered whether their effects on affinity could be rationalized structurally (Fig 4). We

function) automatically prioritizes molecules with the largest expected improvement to binding affinity for each target, regardless of starting seed, so this seed was not prioritized in subsequent rounds of design and no affinity matured designs were generated.

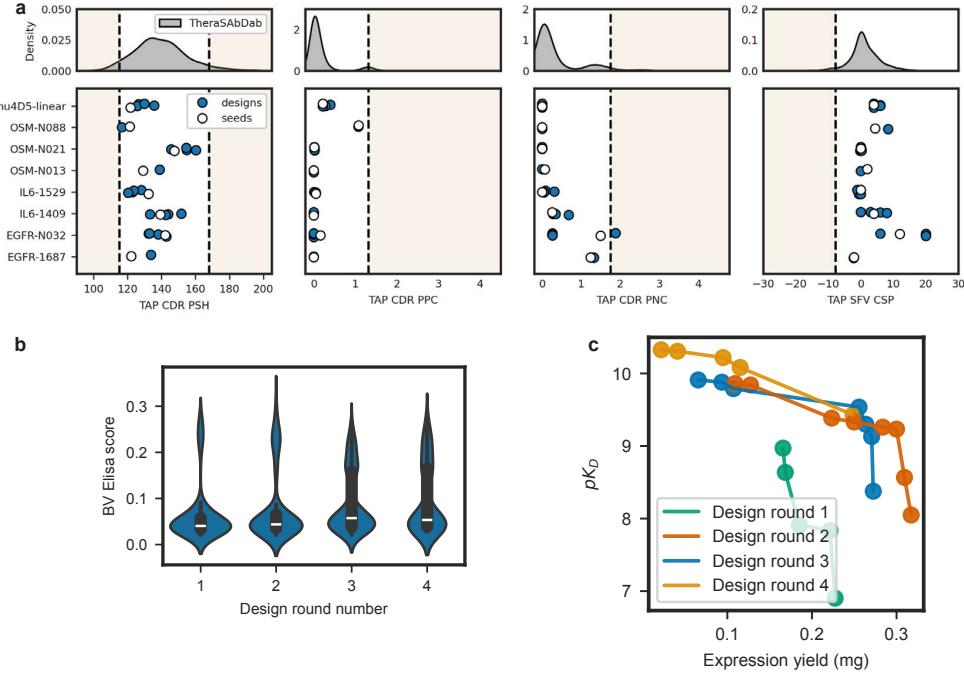


Fig. 3: Multi-property optimization of lead candidates includes improving expression yield and *in silico* developability properties. **a**, All but one of the tightest binding design variants stay in acceptable ranges for developability guidelines given by the Therapeutic Antibody Profiler. **b** Predicted BV Elisa scores, from a surrogate model trained on BV Elisa data, a measure of non-specificity. BVE scores < 1.0 indicate no non-specificity risk. **c**, Pareto frontier of non-dominated designs on the pK_D versus expression yield property landscape for each design round.

first investigated two designs that achieved $>4\times$ improved binding ($\Delta pK_D = 0.688$ (ERBB2-P00873, 5 edits, 1 insertion) and 0.638 (ERBB2-P00965, 8 edits, 1 deletion)) relative to the anti-HER2 antibody Trastuzumab ($pK_D = 8.3$). From 10 independent minimizations of the native crystal structure (PDB ID: 1N8Z), we determined a union set of 17 antigen-contacting residues in the CDR loops through energetic analysis with Rosetta [39, 40] (Fig 4a). Modeling the singular edit at an existing heavy-chain contact (HC:N65Q in CDRH2)—a region with highly dynamic interactions (alternative rotamers in 4/10 simulations for HC:Y67, 3/10 for Ag:E558)—revealed a new hydrogen bond to HC:Y67 that stabilizes a multi-residue chain of interfacial contacts (4a inset), demonstrating that direct optimization of interface positions indeed occurs and may exhibit epistatic effects on binding. However, all other edits made across both designs (Supplementary Figure A10) are distinct from the native contacts, with many outside the CDRs, including an insertion in one design and a deletion in the other. Furthermore, the two designs employ orthogonal edits, only sharing one position with different mutations (HC:S73R and HC:S73A) that is distant from the antigen and thus

280 plausibly a neutral mutation. Folding of both sequences confirmed that the structural
281 integrity of non-CDRH3 loops was maintained despite these changes (Supplementary
282 Figure A10).

283 This interface structure analysis suggests that our sequence-based design meth-
284 ods learn important structural and biophysical priors, resulting in optimized designs
285 that preserve or optimize important energetic contacts necessary for binding, while
286 introducing potentially epistatic mutations that improve binding affinity. Affinity
287 maturation can occur via multiple routes [41] using distinct mutations that are not
288 necessarily physically localized near the interface nor the antigen. We find that inser-
289 tions and deletions are successfully introduced by our models and are tolerated, but
290 not generally necessary for optimization.

291 2.5 Crystal structures of designed antibodies reveal 292 mechanistic roles of mutations

293 To experimentally verify the varied effects of designed mutations across a broader
294 range of targets, we obtained apo crystal structures of four seeds and four designs
295 (Supplementary Table A2, Supplementary Figure A11), in addition to two previously
296 published (9MU1 and 9MSW [32]), spanning three unique targets (Fig 4b). SPR
297 kinetic parameters and affinities are reported for all designs in this figure in Supple-
298 mentary Table A3. For the non-OSM seeds, all crystallized designs are $>10\times$ improved
299 binders. We observed remarkable conservation of overall CDR loop structure in 5/6
300 designs (consistent with predicted structures in Supplementary Figure A2), primarily
301 through newly introduced hydrogen-bonding interactions (Fig 4b insets) that serve to
302 further reinforce the seed loop conformation, which may contribute to affinity improve-
303 ments via structural stabilization. Several other mutations make no new interactions
304 with other residues (e.g. the mutations on CDRH2 of 9MUZ) but are highly solvent
305 exposed and thus may constitute direct contacts to the antigen that are introduced
306 without adversely affecting the native backbone.

307 The exception to this arose in the structural series for seed OSM-N021, where
308 isosteric packing of GY mutations in the stem region of CDRH3 for design 9N7M
309 ($\Delta pK_D = 0.56$, $3.6\times$ improvement) retained the conformations of all CDR loops,
310 yet SY mutations at the same positions in design 9N7O ($\Delta pK_D = 0.65$, nearly $4.5\times$
311 improvement) produced an alternative loop conformation via novel hydrogen-bonding
312 interactions that may contribute to its affinity maturation by more directly exposing
313 the CDRH3 residues. All three OSM seeds proved recalcitrant to optimization (along
314 with hu4D5-linear, IL6-951, EGFR-N065; Fig 2d,f), though we note that these all
315 have starting pK_D values of $>=8.3$ (5 nM) and thus may be close to their local
316 optima. It is therefore plausible that, in contrast to the native-contact-preserving edits
317 to hu4D5-linear (Fig 4a) and the conformation-stabilizing edits to the other seeds
318 (Fig 4b), the structural changes induced by edits in 9N7O may actually be critical to
319 pushing affinity out of a local optimum imposed by the native backbone. The rarity
320 of this occurrence is suggestive that structural consistency imposed by training data
321 and filters (see “Biophysical Priors” in Section 4.1.2 of the Methods) is only sacrificed
322 where more conservative mutations are insufficient.

323 3 Discussion

324 Here, we introduced the “Lab-in-the-loop” (LitL) system for continuous optimization
325 of lead antibody candidates for multiple properties, most notably binding affinity and
326 expression, against 4 distinct antigens of therapeutic interest: EGFR, OSM, HER2,
327 and IL-6. LitL orchestrates generative and discriminative machine learning methods
328 to explore sequence space, while remaining tightly coupled to high-quality, medium-
329 throughput affinity and expression experimental feedback via active learning. Over
330 four experimental iterations, we produced $>3\text{--}100\times$ higher binding affinity antibodies
331 for ten lead molecules to four different antigen targets, while simultaneously retaining
332 or improving upon multiple other essential properties. Apo-state crystal structures
333 of ten lead candidate and affinity matured designs (two previously published [32])
334 unambiguously demonstrated that LitL generates and selects designs possessing ther-
335 apeutically relevant affinities ($>10\times$ improvement) while retaining the seed CDR loop
336 conformations, in part through newly introduced rigidifying mutations and stabiliz-
337 ing intramolecular interactions. The models and filters described herein span both
338 sequence and structural inputs, incorporating domain knowledge, biophysical priors,
339 bioinformatics data, and model-driven *in vitro* data generation. We further demon-
340 strate that sequence-based generative models can produce structurally interpretable
341 mutations when a seed co-complex is available.

342 Because of the complexity of drug discovery, an immensely challenging multi-
343 property optimization problem driven by low-throughput, noisy experimental feed-
344 back, we find that employing machine learning systems and ensembling over compu-
345 tational methods is necessary to mitigate the limitations of any particular approach
346 and address all the relevant challenges. Future work will unite these capabilities and
347 parts of the drug discovery process beyond early stage discovery into an “end-to-end”
348 machine learning system. By demonstrating acceleration of state-of-the-art experi-
349 mental antibody engineering for therapeutic candidate antibodies and tackling the
350 complexity of therapeutic discovery campaigns, we believe that LitL represents a sig-
351 nificant step towards realizing the promise of ML and computation in large molecule
352 drug discovery.

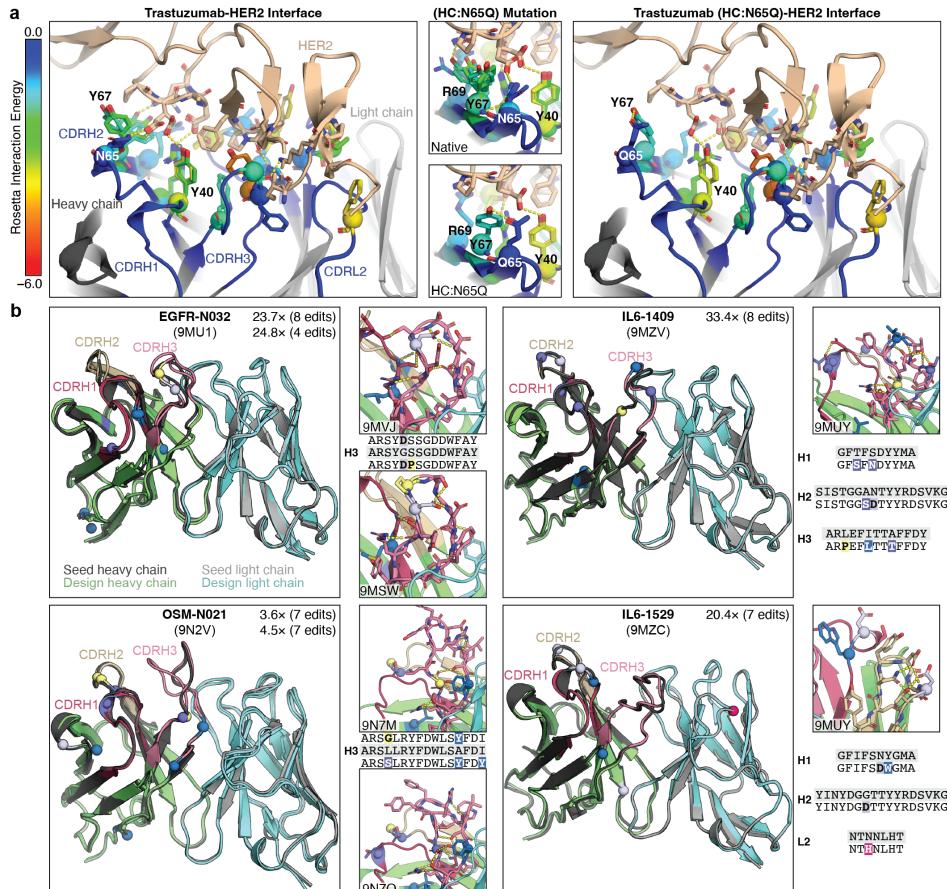


Fig. 4: Interface structure modeling and experimental structure determination yield mechanistic insight into mutational effects. **a**, Native interactions from multiple replicate minimizations of Trastuzumab (hu4D5-linear) are shown as spheres and colored by Rosetta interaction energy (more negative is stronger), with hydrogen bonds depicted as yellow dashes. The CDRH2 interactions are highly dynamic, exhibiting multiple rotameric conformations with variable binding energy. Upon modeling the only mutation at a native contact from two separate $>3\times$ binders (HC:N65Q), a new hydrogen-bond between Q65-Y67 stabilizes the rotamers (highlighted in insets). All other mutations are separate from the native contacts (see Supplementary Figure A10), and also appear several places in the framework region, indicating that they act by indirect effects. **b**, Crystal structures of four seeds and six designs highlight very high structural consistency in almost all cases. For each design, the fold-increase in affinity and number of edits is specified. Edit positions are colored as: nonpolar (blue), polar (purple), negatively charged (lavender), positively charged (magenta), and conformational (yellow). The insets highlight regions of notable new interactions, primarily in CDRH3. Except for 9N7M, at least one edit in each design expands the intramolecular hydrogen-bond network within the CDRs. For the selected CDR sequences reported, the seed sequence is on a grey background and design sequences (insets) have mutations highlighted to match the renders.

353 4 Methods

354 4.1 Computational methods

355 4.1.1 Generative models

356 **Unguided sampling for hit diversification.** Unguided sampling for hit diversification
357 is performed with discrete Walk-Jump Sampling (dWJS) and the Sequence-based
358 Variational Diffusion Model (SeqVDM), introduced in Frey *et al.* [30]. dWJS is a
359 generative model that combines separately trained score- and energy-based models
360 (EBM) to learn a smoothed distribution of *noisy* data, sample noisy data from this
361 smoother space with Langevin Markov Chain Monte Carlo (MCMC) and gradients
362 from the EBM, and then denoise to “clean” data with the score-based model. dWJS
363 builds on the Neural Empirical Bayes [42] formalism and generates 97-100% expressing
364 and up to 70% functional, binding antibodies in *in vitro* experiments [30]. SeqVDM is
365 adapted from [43] for protein sequences; it projects discrete sequences into a continuous
366 latent space and performs denoising diffusion in the latent space. Further details
367 for both dWJS and SeqVDM are provided in [30].

368 **Guided sampling for lead optimization.** Latent Multi-Objective Bayesian
369 Optimization (LaMBO-2), introduced in Gruver & Stanton *et al.* [31], and Property
370 Enhancer (PropEn) [33], are used for lead optimization via guided sampling. LaMBO-
371 2 is a method for controlled categorical diffusion that enables sampling high likelihood,
372 expressing protein sequences that are optimized with respect to an acquisition function,
373 enabling multi-objective design. PropEn is a model designed to implicitly guide
374 design optimization without the need for training a separate discriminator. Instead,
375 PropEn leverages a matched training dataset. Initially, each sequence in the dataset
376 is paired with a similar sequence that exhibits superior binding properties. A Seq2Seq
377 model is then trained to take a lower affinity sequence as input and reconstruct its
378 higher affinity counterpart. DyAb [32] leverages pairwise representations to predict
379 property differences between antibody sequences. Here, it was employed as a ranking
380 model to score combinations of known mutations, and as a generative model when
381 combined with a genetic algorithm. More details regarding DyAb can be found in [32].

382 4.1.2 Biophysical priors

383 Domain knowledge related to antibody engineering and protein design are incor-
384 porated into generative methods in the form of biological priors to restrict the
385 high-dimensional search space and improve the rate of producing designs that bind to
386 the antigen of interest. All the generative models used in this paper are *sequence-based*
387 models; they are trained on amino acid sequences and emit amino acid sequences.
388 While no protein structures are explicitly generated by our methods, protein structure
389 information is included via strong biological priors. Models are trained on AHo-aligned
390 [44] antibody sequences, a scheme with particular emphasis on structural consis-
391 tency. AHo alignment introduces alignment tokens (“-”) and ensures a global sequence
392 alignment across all antibodies, clearly separating an antibody sequence into distinct
393 structural elements (framework and complementarity determining regions (CDRs)).

394 Training on aligned sequences [30, 31, 45, 46] is an established method for improving
395 the performance of protein generative models and implicitly providing a strong
396 structural prior. AHo alignment also enables our methods to easily introduce length
397 changes (insertions or deletions), by sampling alignment tokens. However, at sampling
398 time we often mask out the alignment token, preventing length changes, which may
399 attenuate binding. Similarly, some designs restrict mutations to only the CDRs or
400 subsets of the CDRs, which are generally thought to be most important for binding,
401 though it is known that framework positions can also dramatically impact affinity,
402 both individually [47, 48] and as part of humanization[49]. Our biophysical priors and
403 design methods are built to preserve the binding mode of the seed antibody [50].

404 **4.1.3 Property prediction models**

405 **Multi-task fine-tuning.** For each LitL round, denoted by t , we first perform *dis-*
406 *criminative model selection*, using the generalization error on the last round ($t - 1$) to
407 estimate the generalization error on the next round. A uniform random holdout set
408 can be used as the validation set at $t = 0$. The best performing oracle is used for each
409 property, as measured by a pre-determined metric on the hold-out set. Framing model
410 selection as an open competition encourages rapid development of surrogate models
411 and ensures that a diverse range of ML approaches are tried for challenging biological
412 properties. An exceptionally performant model emerged during LitL competition:
413 Cortex [31], a multi-task fine-tuning framework that uses pre-trained Language models
414 for Biological Sequence Transformation and Evolutionary Representation (LBSTER)
415 [51] to simultaneously model all properties of interest.

416 **Pre-trained representation learning.** Language models for Biological
417 Sequence Transformation and Evolutionary Representation (LBSTER) is used for
418 pre-training representation learning models for multi-task property prediction and
419 optimization with Cortex. LBSTER includes both Masked Language Models (MLMs)
420 based on the ESM-2 model architecture [52] and Causal Language Models (CLMs)
421 based on the Llama-2 model architecture [53]. LBSTER models are trained on
422 Uniref50 [54] and the Observed Antibody Space [55], de-duplicated with MMSeqs2
423 [56]. We use the protein language model cramming strategy introduced in Frey *et al.*
424 [51] to pre-train performative LBSTER models in only a few GPU days.

425 **4.1.4 Out-of-distribution detection**

426 OOD detection is performed on designed sequences to determine if they are in-
427 distribution for property prediction models and therefore can be reliably ranked by
428 the ranking model. OOD detection is implemented with the distributional confor-
429 mity score (DCS), as introduced in Frey *et al.* [30], which is a measure of how likely
430 generated samples are with respect to a reference set of already known binders or
431 expressors. DCS uses a likelihood under a joint density of statistical properties, includ-
432 ing log-probability under a protein language model, and sequence-based properties
433 like hydrophobicity and molecular weight, calculated with BioPython [57]. Kernel
434 density estimation (KDE) is used to compute the joint density, although any density
435 estimator may be used.

436 4.1.5 Filters and quality control

437 Designs are annotated to ensure that they do not introduce chemical liabilities (e.g.,
438 glycosylation and deamidation motifs) and that conserved amino acids in antibody
439 sequences are preserved (e.g., canonical cysteines). Additional quality control checks
440 are performed to prevent germline reversion (which may lead to affinity attenuation).
441 Further constraints are imposed across all designs and specific generative models to
442 encourage generative models to explore local neighborhoods in sequence space around
443 lead antibodies, so that designs are more likely to be in-distribution for property pre-
444 diction models. In the first round of design, a maximum edit distance cap of 6 edits
445 from the lead is enforced. In the second round, this cap is increased to 8, and in the
446 third round, it is increased to 12. Some generative models restrict edits to partic-
447 ular regions (e.g., dWJS only introduces mutations in CDRs), while other methods
448 including PropEn and LaMBO-2 introduce mutations across the entire sequence.

449 4.1.6 Ranking and active learning

450 We use Noisy Expected Hypervolume Improvement (NEHVI) [34] as our acquisition
451 function to optimize the selection of designs with respect to multiple properties $\mathbf{x} \in$
452 \mathcal{X} that we represent as a vector-valued function $f : \mathcal{X} \rightarrow \mathbb{R}^m$. Because this is a
453 multi-objective optimization problem, we are looking for *non-dominated* (i.e., Pareto
454 optimal) designs. A design, x_1 , dominates another design, x_2 , if it no worse than x_2
455 for all properties and strictly better for at least one property. We seek to produce and
456 select a set of non-dominated designs to assay in the lab in each successive iteration
457 of LitL. In general, the properties we care about are *black-box* objectives, with no
458 closed-form expression, so we learn about these distributions via expensive, slow, noisy
459 measurements.

To select a “batch” of designs, we obtain a posterior distribution $p(f|D)$ over functions for real properties, given a dataset, D , of measurements. We choose a reference design x_{ref} as a lower bound and never choose any design that is predicted to be worse than x_{ref} for any property. When considering whether a design should be added to our batch, we compute its “hypervolume improvement” (HVI), which is the difference in hypervolumes (in property space) between the design and the designs already in our batch. Finding the subset of \mathcal{X} that maximizes HVI is an NP-hard problem, but a greedy algorithm works quite well. To account for epistemic uncertainty, we rely on surrogate models and marginalize f , yielding the noisy expected HVI (NEHVI) for a design x' ,

$$\text{NEHVI}(\mathbf{x}') = \mathbb{E}_{f|D}[\text{HVI}(\mathbf{x}')]. \quad (2)$$

460 4.1.7 Interpretability

461 Sequence-level interpretability is performed with Cortex, as described in Adebayo *et*
462 *al.* [58]. Independent Gaussian noise corruption is introduced as a computationally
463 efficient alternative to adversarial training, without deteriorating classifier perfor-
464 mance. Gaussian corruption is shown to confer faithfulness to feature attributions,
465 i.e., saliency maps computed with classifiers that are regularized during training with

466 Gaussian corruption correctly ignores spurious distractor features. This regularization
467 allows inspection of gradient-based saliency maps to determine the residues and
468 mutations that classifier predictions are most sensitive to.

469 4.1.8 Rosetta biomolecular modeling

470 Input structures were prepared using Rosetta [39] gradient descent minimization fol-
471 lowed by FastRelax [59, 60] with the *beta_nov16* scorefunction [40] and the following
472 flags:

```
473 -corrections:beta_nov16 -ex1 -ex2 -use_input_sc -flip_HNQ -no_optH  
474 false -packing:repack_only true -relax:constrain_relax_to_start_coords  
475 -relax:ramp_constraints false
```

476 In the case of the seed hu4D5-linear (Trastuzumab), coordinates were obtained from
477 PDB code 1n8z. The HC:N65Q mutation was introduced using the MutateResidue
478 mover. The entire structure was first minimized using MinMover then the interface
479 was allowed to repack using FastRelax (InterfaceRelax2019 script) with nonramping
480 backbone constraints to the starting coordinates applied during both steps. This
481 MinMover+FastRelax protocol was found to produce plausible local variation in
482 sampling more effectively than either Mover alone, while still yielding low-energy
483 structures expected of a high-affinity protein complex. Two-body interaction ener-
484 gies were computed using the InteractionEnergyMetric, which includes through-space
485 terms including dispersion, solvation, electrostatic, and hydrogen bonding interac-
486 tions. Contacts were defined as any residue that has ≤ -0.6 REU interaction energy
487 to at least one individual residue on the antigen side. Per-residue values were writ-
488 ten to the B-factor columns and used for visualization. All other visualized structures
489 were folded directly from the seed or design sequences using ABodyBuilder2 and saved
490 with refinement.

491 4.2 Engineering infrastructure

492 4.2.1 Data engineering

493 Our data backend serves as an intermediary connecting functional assay data curated
494 by scientists, direct instrument access, as well as data drawn from public and propri-
495 etary databases. In the context of LitL, this constitutes the retrieval and processing
496 of assay data to enable retraining and validation of our models. At a high level, we
497 implemented abstractions for periodic (i) access, (ii) retrieval, and (iii) transforma-
498 tion of data to produce and update core datasets. Together this enables database agnostic
499 deployable pipeline orchestration that persists dataset artifacts in the cloud, provid-
500 ing a common setting in which data from disparate sources may be jointly queried,
501 combined, quality controlled, and ultimately packaged into ML-ready datasets.

502 In addition to basic data integrity, reproducibility, and versioning carried out auto-
503 matically by data pipelines, quality control standardizes assay registration and data
504 capture. We set protocols for instrument runs, machine data capture, as well as the
505 application of human and heuristic annotation. As a result of these efforts, we have
506 dramatically increased available data across many more features and with sufficient
507 assay context to drive model improvement.

508 Designed sequence libraries are verified and registered in the internal large molecule
509 registration system, and each sequence receives a unique identifier. Protein production
510 requests at a scale of 1 ml are submitted via an internal protein production system,
511 and all related request and fulfillment metadata are collected into databases and linked
512 with unique identifiers from the registration system. Subsequent functional assays are
513 recorded in a structured fashion via an internal Lab Information Management System
514 (LIMS) that also supports plating provenance and registrations across all manual and
515 automated Antibody Engineering workflows. Data for each molecule is centralized and
516 integrated into a single repository (Supplementary Figure A9) and made available for
517 consumption by downstream machine learning workflows.

518 **4.3 Design submission**

519 When a design is submitted, metrics to quantify binding affinity and expression yield
520 predictions are computed. Litl is orchestrated using a Kubernetes-based computa-
521 tion pipeline, providing a structured and modular approach to pipeline development.
522 The frontend user interface manages instructions for file submission, file upload to
523 persistent storage, and visualizations of the uploaded antibody data. Alternatively,
524 a Python client can be used for programmatic submission and retrieval of pipeline
525 results. Following file upload, the server performs data formatting checks and trig-
526 gers the computation pipeline, which executes antibody registration, quality checks,
527 enumeration, folding, inference with property prediction models, out-of-distribution
528 checks, and quality control filtering. Parallel execution is employed for independent
529 tasks. Processed data undergo additional quality control before being passed to the
530 ranking pipeline, using an active learning framework. The selection of the best can-
531 didate designs is then passed into an automated visualization tool and a formatting
532 step for DNA synthesis order.

533 **4.4 Experimental methods**

534 Experimental protocols follow Hsiao *et al.* [4, 61] and Wu *et al.* [35]. The methods are
535 described here for convenience.

536 **4.4.1 Immunization**

537 We follow the animal immunization procedure described in Hsiao *et al.* [4]. Animals
538 were immunized and both splenocytes and lymph node cells were used for hybridoma
539 fusion. RNeasy Mini Kit (Qiagen) was used for total RNA extraction from hybridoma
540 cells. Reverse-transcription (RT) PCR of heavy and light chains was done with a
541 SMARTer RACE cDNA Amplification Kit (Clontech). The forward Universal Primer
542 Mix from SMARTer RACE cDNA Amplification Kit was used for the PCR step. The
543 dideoxy sequencing method was used to directly sequence amplified PCR products.

544 **4.4.2 Repertoire sequencing and mining**

545 Direct lysis with ammonium chloride was used to remove red blood cells. Spleen and
546 bone marrow cells were also collected. Buffer RLT (Qiagen) was used to lyse remain-
547 ing splenocytes and bone marrow cells. The RNeasy Mini Kit was used to purify RNA

according to the instructions from the manufacturer. Amplicons were purified from agarose gels and sequenced with an Illumina MiSeq instrument. Sequences were processed to 1) extract sequences within variable region boundaries, 2) remove incomplete sequences and sequences with ambiguous residues, 3) remove the first eight residues of framework region 1 (FR1) encoded by amplification primers, 4) extract sequences matching mAb CDR3 length and identity threshold of 57% for CDR H3 sequences, and 5) identify and count unique variable region sequences [4].

To compare against design libraries from generative models, some heavy chain variants of seeds (with light chain fixed) were either chosen directly from the bulk NGS data or designed based on observed patterns in the neighborhood of the seed sequences. First, we collected all sequences that have the same lengths as seeds and are within 12 Hamming distance over the entire variable region and 3 Hamming distance in CDR-H3 from the seeds. Among these, we picked several observed sequences that are most distant by Hamming distance from inferred germline sequences associated with seed sequences. Next, we used the collected sequences to construct Position Specific Scoring Matrix (PSSM) including each seed and their neighbors. We then designed one sequence per seed by sampling the frequency mode of the PSSM at each position while preventing reversion to inferred germline sequence residues associated with the seed. We also included observed sequences that rank high according to the PSSM (i.e. sum of frequencies of observed residues).

4.4.3 Plasmid construction and antibody production

All antibodies were expressed utilizing GLFs as previously described [35]. Briefly, VH and VL fragments were synthesized with 40bp overhang sequences on both sides (Twist Bioscience) and subsequently integrated via Gibson assembly into in-house designed pRK5 expression vectors containing the appropriate heavy chain or light chain constant regions and other elements required for expression (promoter, ORF, poly-A tail) [62]. The product of this reaction was used as the template for GLF amplification, with primers located back-to-back in the noncoding region of the vector, amplifying the whole vector-sized linear DNA insert. This insert contains 2kb extra sequence on each side of the expression cassette, offering extra protection from degradation.

1mL scale IgG expressions were transiently conducted in Expi293F cells (ThermoFisher Scientific). Cells were transfected using PEIpro (PolyPlus PEIpro, Reference #: 115-01L. 1mg/ml). 1 μ L PEIpro was diluted into 100 μ L serum-free medium, then 6 μ L of GLFs (1:1 mixture of heavy chain and light chain, approximate concentration 200 μ g/ μ L) were added and incubated for 10 min. The PEI-DNA complex was added to the 0.85ml cells, and cells were cultured in a 96-deep well plate at 37 °C and 8% CO₂ for 7 days, with 150 μ L fresh medium added 24hr post transfection.

The supernatant was collected and purified with a 1-step purification using Protein A affinity resin (MabSelect SuRe™, Cytiva). Quality control of antibody purity was determined by SDS-PAGE with Coomassie blue staining and A280 absorbance to measure protein concentration.

590 **4.4.4 SPR affinity characterization**

591 Surface Plasmon Resonance (SPR) was performed on a Biacore 8K+ (Cytiva). The
592 instrument was primed with HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl,
593 3 mM EDTA, 0.005% v/v Surfactant P20), and antigens or antibodies were also
594 diluted into the same buffer for all SPR experiments. A Protein A chip was used
595 for antibody capture in all experiments. Experiments were either done using single
596 cycle kinetics (the antigen series successively injected from low to high concentrations,
597 followed by one dissociation step) or multi cycle kinetics (each cycle using one antigen
598 concentration for association followed by a dissociation step, and then repeated for
599 each antigen concentration). In a typical experiment, five concentrations (0.16nM,
600 0.8nM, 4nM, 20nM and 100nM) for 180s association time, 600s dissociation time with
601 a Flow rate of 30 μ L/min. All data were analyzed using Biacore Insight software using
602 a 1:1 binding model.

603 **4.4.5 BV ELISA**

604 BV ELISA was performed following [38], using chimeric mouse/human IgG1 anti-
605 bodies. Antibodies were purified by protein A affinity chromatography and SEC. BV
606 ELISA is an assay based on ELISA detection of non-specific binders to baculovirus
607 (BV) particles. BV ELISA identifies antibodies that have an increased risk for poor
608 pharmacokinetic properties (fast clearance). Molecules with a BV ELISA score > 1.0
609 indicate an increased risk. We train a surrogate model by fine-tuning a pre-trained
610 LBSTER [51] protein language model on an internally generated dataset of 2,305 BV
611 ELISA measurements.

612 **4.4.6 Protein expression, purification, and structure determination**

613 The Fab constructs were transfected into CHO cells at a 1:2 (HC:LC) DNA ratio, and
614 expressed for 10 days. After harvesting, the supernatant was collected for purification
615 using GammaBind Plus Sepharose (Cytiva) followed by size exclusion or SP cation
616 exchange chromatography. The purified Fabs were formulated in 20mM histidine
617 acetate pH 5.5, 150mM sodium chloride.

618 The Fabs were concentrated to 10mg/mL and crystallized by vapor diffusion as
619 follows: 9MVJ: 2.0 M Ammonium sulfate, 0.2 M Sodium chloride, 0.1 M Sodium
620 cacodylate pH 6.5; 9MUZ: 0.2 M Sodium Potassium phosphate, 20% v/v Polyethylene
621 glycol 3350; 9MUY: 0.1 M Ammonium acetate, 0.1 M Zinc chloride, 0.1 M Bis-Tris
622 pH 7.2, 15% v/v PEG Smear High; 9MZC: 0.09 M NPS, 0.1 M BS1 pH6.5, 50% v/v,
623 PM2 (Polymer); 9MZV: 0.2 M Calcium chloride, 0.1 M Tris-HCl pH 8.5, 20% w/v
624 PEG 4000; 9N2V: 0.2 M Sodium chloride, 0.1 M Imidazole pH 8.0, 0.4 M NaH₂PO₄,
625 1.6 M K₂HPO₄; 9N7M: 0.2 M Ammonium sulfate, 0.1M MES pH 6.5, 30% w/v PEG
626 5000 MME; 9N7O: 0.1 M Bis-Tris propane pH 9.0, 10% v/v PEG 200, 18% w/v PEG
627 8000;

628 Crystals were cryoprotected by addition of ethylene glycol. Data were collected at
629 synchrotron sources, as indicated in Supplementary Table A2, and data were scaled
630 using XDS [63]. The structures were determined by molecular replacement, using
631 PHASER [64], and the models were built in COOT [65] and subsequently refined

632 in PHENIX [66] to final statistics presented in Supplementary Table A2. The sim-
633 ulated annealing omit difference electron density maps (Supplementary Figure A11)
634 demonstrate the fidelity of the final models with the electron density.

635 **Supplementary information** Supplementary files accompany this article.

636 **Acknowledgments** We are grateful to Viva Biotech staff, Yanshu Dou, Chunheng
637 Huo, Yu Tang for contributions to the crystal structures and to Feng Wang and
638 Yoana Dimitrova, for coordinating structural experiments. The authors would like to
639 thank the facilities and staff of Diamond Light Source beamline I03 and Shanghai
640 Synchrotron Radiation Facility (SSRF) beamline BL02U1 for contributions towards
641 crystallographic data collection. We thank the experimental facility and the technical
642 services provided by “The Synchrotron Radiation Protein Crystallography Core Facil-
643 ity (SPXF) of the National Core Facility for Biopharmaceuticals (NCFB), National
644 Science and Technology Council (NSTC)” and the “National Synchrotron Radiation
645 Research Center (NSRRC)”, a national user facility supported by the National Sci-
646 ence and Technology Council of Taiwan (NSTC), Taiwan (R.O.C.). We acknowledge
647 DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the
648 provision of experimental facilities. Parts of this research were carried out at beamline
649 P11.

650 We thank the Genentech NGS core lab for the repertoire sequence data used in
651 this study.

652 The authors acknowledge the entire Prescient Design team and the Antibody Engi-
653 neering department at Genentech for providing helpful discussions and input that
654 contributed to the research results reported within this paper. The authors would espe-
655 cially like to acknowledge Candice Wertz for coordination and project management
656 for manuscript preparation.

657 **Declarations**

658 **Funding**

659 All work described in this paper was funded by Genentech Inc., South San Francisco,
660 CA.

661 **Competing interests**

662 All authors are or were employees of Genentech Inc. (a member of the Roche Group)
663 or Roche, and may hold Roche stock or related interests.

664 **Ethics approval**

665 All animal studies were performed in animal facilities accredited by the Association
666 for Assessment and Accreditation of Laboratory Animal Care International. The pro-
667 cedures for animal studies were compliant under the Institutional Animal Care and
668 Use Committee of the facility.

669 **Availability of data and materials**

670 LBSTER models were trained using the [UniRef50](#), [PDB](#), and [Observed Antibody](#)
671 [Space](#) databases. All other models were trained on internally generated datasets
672 described in the experimental methods, Section 4.4.

673 Crystallographic coordinates and data were deposited at the RCSB with accession
674 codes 9MUY, 9MUZ, 9MVJ, 9MZC, 9MZV, 9N2V, 9N7M, 9N7O.

675 **Code availability**

676 Discrete Walk-Jump Sampling, PropEn, LBSTER, and Cortex are available on
677 the [Prescient Design GitHub](#). DyAb is available as a model within LBSTER. The
678 Prescient standard library (Beignet) and LaMBO-2 are available on GitHub.

679 **Authors' contributions**

680 R.G.A., D.B., R.B., K.C., H.D., N.C.F., V.G., I.H., R.K., S.K., J.R.K., J.T.K., J.L.V.,
681 J.H.L., A.L., S.R., F.S., S.D.S., and A.M.W. conceived of and designed the work. R.A.,
682 J.B., T.B., P.C., T.D., H.D., A.D., J.F., N.C.F., V.G., A.G., J.H., H.I., A.I., S.J., S.K.,
683 M.K., J.K., J.L.V., A.L.F., E.L., W.C.L., J.Y.Y.L., S.P.M., E.M., S.N., J.W.P., J.P.,
684 S.R., F.S., S.D.S., N.T., H.T., A.W., A.M.W., B.W., S.W., and K.Z. acquired the
685 data. R.A., D.B., J.B., T.B., P.C., A.C., T.D., H.D., N.C.F., V.G., A.G., I.H., H.I.,
686 T.J., S.K., J.R.K., J.K., J.L.F., J.H.L., E.L., D.L., W.C.L., A.L., J.L., O.M., E.M.,
687 K.M., H.M.P., J.P., S.R., F.S., I.S., A.S., S.D.S., E.W., A.W., A.M.W., B.W.,
688 and K.Z. analyzed and interpreted the data. R.A., D.B., T.B., H.D., N.C.F., V.G.,
689 A.G., J.H., H.I., A.I., S.K., M.K., J.K., A.L.F., E.L., J.Y.Y.L., A.L., J.L., S.P.M.,
690 O.M., S.N., J.W.P., J.P., S.R., S.S., S.D.S., N.T., E.W., A.M.W., and K.Z. created
691 methods and algorithms presented in this work. R.A., R.B., K.C., A.D., J.F., N.C.F.,
692 V.G., J.H., I.H., S.J., T.J., R.K., S.K., J.R.K., J.T.K., J.L.F., J.H.L., A.L., E.M., J.M.,
693 K.M., H.M.P., J.W.P., S.R., A.R., S.R., S.S., I.S., S.D.S., N.T., A.W., A.M.W., and
694 Y.W. wrote and edited the manuscript. R.B., P.C., K.C., H.D., V.G., I.H., R.K., S.K.,
695 J.R.K., J.T.K., J.L.V., A.L., J.M., S.R., A.R., F.S., A.M.W., B.W., and Y.W. super-
696 vised the research. All authors discussed the results and approved the final version of
697 the manuscript.

698

699 Appendix

700 A Supplementary data

Table A1: SPR binding kinetics for select antibody designs reported in the main paper. Full antibody and antigen sequences and additional data, including goodness of fit metrics, are available in the Extended Data.

Antibody Alias	k_a (1/Ms)	k_d (1/s)	K_D (M)	pKD	Target
EGFR-P01451	5.29e+06	4.50e-04	8.50e-11	1.01e+01	EGFR
EGFR-P01484	4.12e+06	6.16e-04	1.50e-10	9.82e+00	EGFR
EGFR-P01503	3.99e+06	6.06e-04	1.52e-10	9.82e+00	EGFR
EGFR-P01519	2.16e+06	1.44e-02	6.68e-09	8.18e+00	EGFR
EGFR-P01606	3.64e+06	5.96e-04	1.64e-10	9.79e+00	EGFR
ERBB2-P00873	1.20e+05	2.01e-04	1.67e-09	8.78e+00	ERBB2
ERBB2-P00961	1.25e+05	3.14e-04	2.52e-09	8.60e+00	ERBB2
ERBB2-P00965	1.22e+05	2.68e-04	2.20e-09	8.66e+00	ERBB2
IL6-P01545	1.27e+06	1.17e-03	9.18e-10	9.04e+00	IL6
IL6-P01548	8.70e+05	8.18e-04	9.41e-10	9.03e+00	IL6
IL6-P01573	2.20e+06	1.19e-03	5.40e-10	9.27e+00	IL6
IL6-P01577	1.13e+06	1.50e-03	1.33e-09	8.88e+00	IL6
IL6-P01615	6.59e+05	9.64e-04	1.46e-09	8.84e+00	IL6
IL6-P01634	5.82e+05	3.53e-04	6.07e-10	9.22e+00	IL6
IL6-P01662	9.76e+05	1.37e-03	1.41e-09	8.85e+00	IL6
IL6-P01770	1.16e+06	1.11e-03	9.61e-10	9.02e+00	IL6
OSM-P01123	1.13e+06	1.83e-03	1.61e-09	8.79e+00	OSM
OSM-P01250	6.64e+05	3.55e-04	5.34e-10	9.27e+00	OSM
OSM-P01344	3.67e+05	3.55e-04	9.67e-10	9.01e+00	OSM
OSM-P01377	1.04e+06	1.72e-03	1.66e-09	8.78e+00	OSM
OSM-P01473	3.31e+05	2.69e-04	8.12e-10	9.09e+00	OSM
OSM-P01508	1.30e+05	3.32e-04	2.55e-09	8.59e+00	OSM
OSM-P01568	8.67e+04	6.60e-04	7.61e-09	8.12e+00	OSM

Table A2: Crystallographic data

PDB code	9MVJ	9MUZ	9MUY	9MZC
X-ray source	NSRRC TPS 05A	SSRF BL02U1	Diamond I03	SSRF BL02U1
Wavelength (Å)	0.99987	0.97918	0.97625	0.97918
Detector	Rayonix MX300HS	Eiger2 S 9M	Eiger2 XE 16M	Eiger2 S 9M
Resolution range (Å)	49.3 – 1.95	147.3 – 1.71	50.9 – 1.97	47.5 – 1.58
Highest res. bin (Å)	2.00 – 1.95	1.74 – 1.71	2.02 – 1.97	1.60 – 1.58
Space group	H ₃ 2	P2 ₁ 2 ₁ 2	C ₂	P2 ₁ 2 ₁ 2
Multiplicity	20.7, 21.2	7.0, 6.0	6.9, 6.6	6.9, 7.1
Complete (%)	100, 100	100, 100	99.9, 99.8	98.1, 95.1
Mean I/σ _I	21.0, 2.5	11.3, 1.9	13.3, 2.2	12.4, 2.0
Wilson B (Å ²)	33.8	19.6	32.1	3.5
CC ½ (highest bin)	0.86	0.69	0.87	0.69
Rmerge (%)	10.7, 159	10.8, 86.1	7.7, 80.9	16.7, 107
# reflections (Rfree set)	59,055 (5,735)	91,776 (8,797)	38,938 (3,770)	62,385 (3,054)
Resolution range (Å)	39.7 – 1.95	73.7 – 1.71	33.3 – 1.97	47.3 – 1.58
Rwork, Rfree (%)	18.0, 19.8	16.6, 19.6	20.8, 22.3	17.5, 20.2
# non-H atoms	3,715	7,682	3,585	4,264
# solvent molecules	328	939	225	606
Rmsd bond lengths (Å)	0.004	0.004	0.005	0.006
Rmsd bond angles (°)	0.668	0.644	0.832	0.913
Ramachandran fav. (%)	97.4	98.1	96.5	97.9
Ramachand. outlier (%)	0	0	0.24	0
Ave B-factor (Å ²)	48.2	23.3	53.9	12.6
Molprobitiy clash score	1.82	1.43	3.97	1.67
PDB code	9MZV	9N2V	9N7M	9N7O
X-ray source	PETRAIII P11	PETRAIII P11	SSRF BL02U1	SSRF BL02U1
Wavelength (Å)	1.03321	1.03321	0.97918	0.97918
Detector	EIGER2 X 16M	EIGER2 X 16M	Eiger2 S 9M	Eiger2 S 9M
Resolution range (Å)	48.3 – 1.90	46.55 – 1.97	46.53 – 2.15	44.9 – 1.48
Highest res. bin (Å)	1.94 – 1.90	2.00 – 1.97	2.20 – 2.15	1.50 – 1.48
Space group	P2 ₁	P4 ₁	P4 ₁	P4 ₁ 2 ₁ 2
Multiplicity	6.9, 6.9	13.7, 13.4	5.1, 4.2	26.1, 21.1
Complete (%)	99.3, 99.1	99.5, 100	99.9, 100	99.8, 100
Mean I/σ _I	12.4, 2.2	16.1, 2.1	7.7, 2.0	13.0, 2.0
Wilson B (Å ²)	20.6	27.8	19.0	19.6
CC ½ (highest bin)	0.68	0.745	0.154	0.306
Rmerge (%)	13.4, 103	15.4, 162	22.4, 70.7	18.2, 184
# reflections (Rfree set)	71,330 (3,484)	95,351 (4,880)	73,985 (3,682)	94,027 (4,649)
Resolution range (Å)	48.3 – 1.90	46.6 – 1.97	41.3 – 2.15	31.8 – 1.48
Rwork, Rfree (%)	17.7, 21.4	17.4, 20.2	21.0, 24.2	17.2, 20.5
# non-H atoms	7,507	7,635	7,474	3,500
# solvent molecules	839	853	826	601
Rmsd bond lengths (Å)	0.002	0.002	0.002	0.006
Rmsd bond angles (°)	0.591	0.589	0.512	0.841
Ramachandran fav. (%)	98.0	98.0	98.1	97.7
Ramachand. outlier (%)	0	0	0	0
Ave B-factor (Å ²)	24.0	38.0	27.5	26.1
Molprobitiy clash score	0.92	0.82	1.15	1.29

Table A3: SPR binding kinetics for designs reported in Figure 4. Full antibody and antigen sequences and additional data, including goodness of fit metrics, are available in the Extended Data. ΔpKD values are reported for measurements where design and seed were on the same plate, leading to two slightly different affinities for the OSM-N021 seed. “N/A” PDBIDs are the two folded anti-HER2 designs for which experimental structures were not obtained. Affinities for 9MSW and 9MU1 were previously reported in [32].

PDBID	Antibody Alias	k_a (1/Ms)	k_d (1/s)	K_D (M)	pKD	Seed pKD	Δ pKD
9MVJ	EGFR-P01451	4.86E+06	6.24E-04	1.28E-10	9.891	8.517	1.374
9MSW	EGFR-P01550	4.60E+06	5.65E-04	1.23E-10	9.911	8.517	1.394
9N7M	OSM-P01295	2.78E+05	1.30E-04	4.68E-10	9.330	8.770	0.560
9N7O	OSM-P01513	3.26E+05	1.41E-04	4.33E-10	9.364	8.708	0.656
9MUZ	IL6-P01770	1.67E+06	1.03E-03	6.17E-10	9.210	7.686	1.524
9MUY	IL6-P01634	1.14E+06	3.03E-04	2.66E-10	9.575	8.266	1.309
N/A	ERBB2-P00873	1.28E+05	1.67E-04	1.30E-09	8.884	8.195	0.689
N/A	ERBB2-P00965	1.39E+05	2.04E-04	1.47E-09	8.833	8.195	0.638

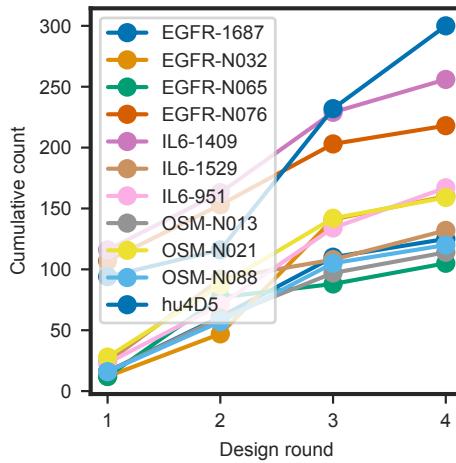


Fig. A1: Cumulative number of labeled (expression yield and binding affinity) designs in the neighborhood of each seed (≤ 16 edit distance) over rounds

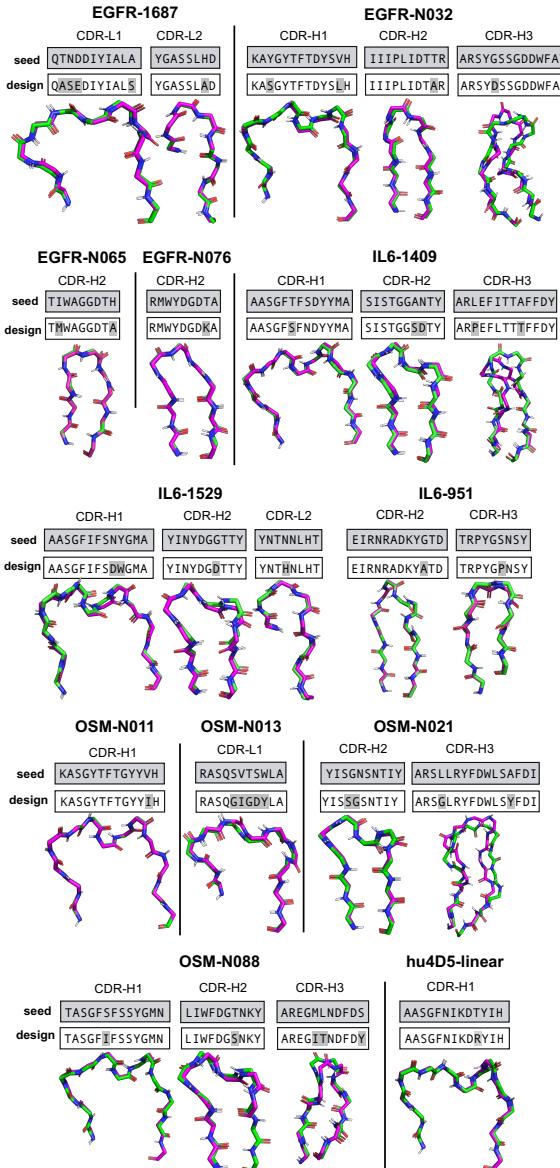


Fig. A2: CDR loop comparison between seeds and designs. Global alignment of seed (green) and affinity-matured designs (magenta). All CDR loops (AHo definitions [44]) containing mutations between seed and design are shown (other mutations may be present in non-CDR regions).

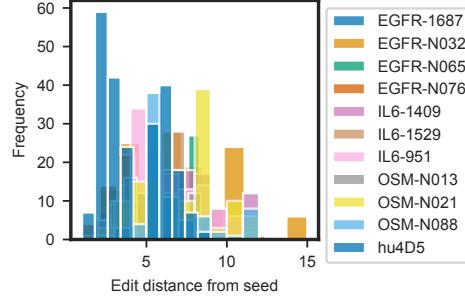


Fig. A3: Histogram of edit distances of all binding designs with respect to their respective seed sequences. Edit distances range from 1 to 14.

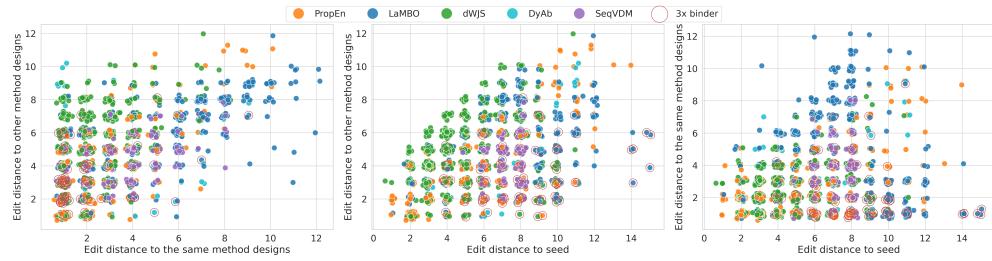


Fig. A4: Edit distances of generated designs to seed, other method designs, and other designs from the same method. It can be seen that 3x binders usually have lower edit distance to both, other method designs and other designs from the same method, but higher edit distance to seed. This highlights the importance of our multi-round setup, that allows us to drift further away from the starting seed. It is also worth noting, that while most methods have rather linear relationship between distance to seed and distance to other method designs, with a slightly higher similarity to other method designs, dWJS can produce tangent designs as shown by low edit distance to seed and high edit distance to designs from other methods.

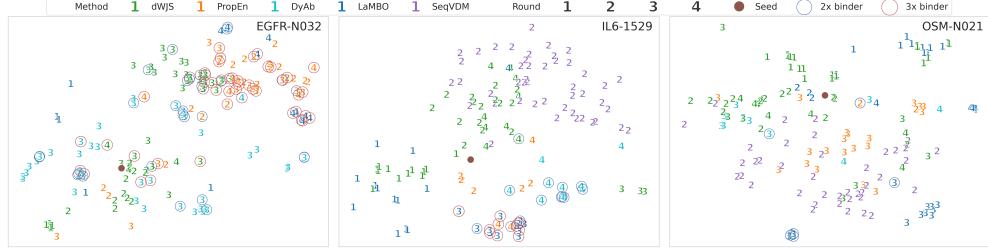


Fig. A5: t-SNE[67] plot of ESM-2 [52] embeddings of designs around 3 selected seeds. We can see that dWJS and SeqVDM, as expected, explore the design space more widely, while designs of PropEn and LaMBO are more localized, especially within the same round. The importance of thoroughly exploring the space around the seed is also highlighted by the fact, that most of the time 2x and 3x binders are tightly grouped and can be far from the seed.

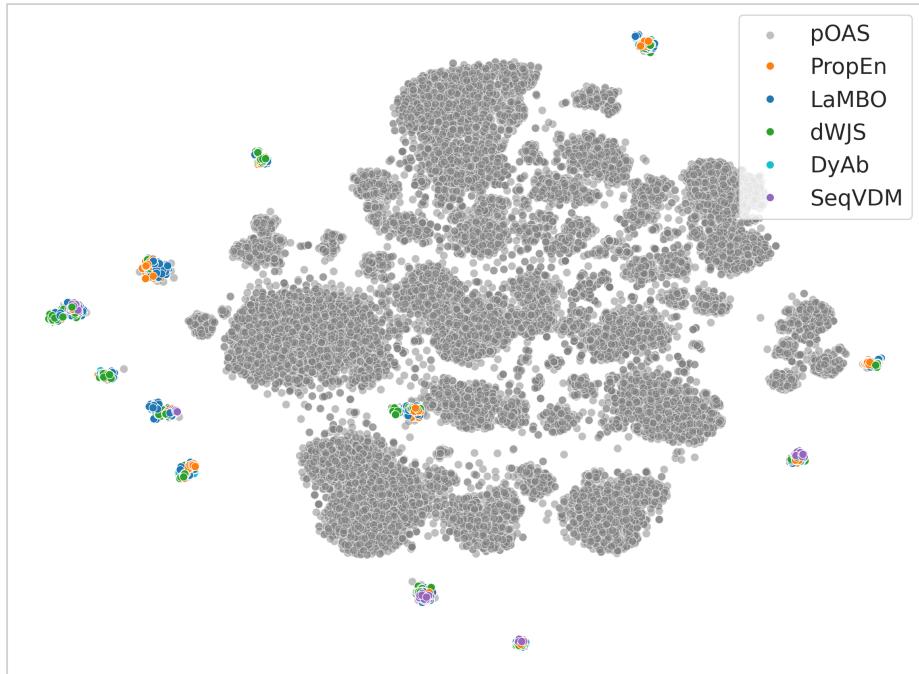


Fig. A6: t-SNE[67] plot of ESM-2 [52] embeddings of designs for all seeds plotted against paired OAS [55] (30k random sub-sample). Our seeds are mainly either on the fringes of this reference distribution or are completely OOD.

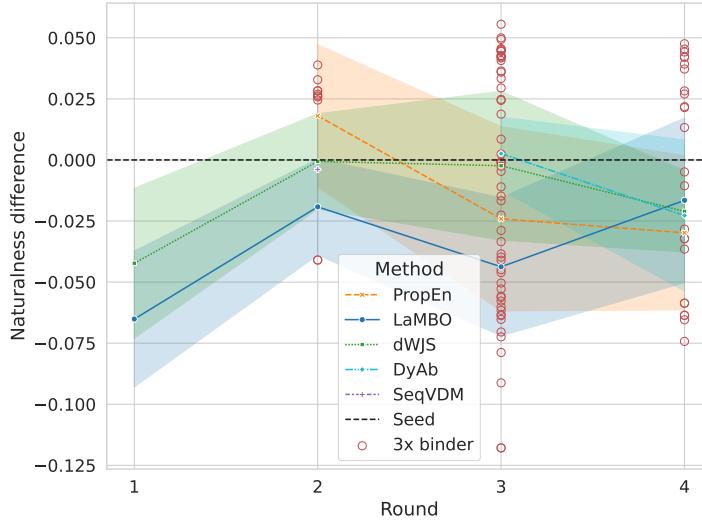


Fig. A7: Naturalness score [68] (inverse perplexity of AntiBERTy PLM [19]) of the designs over the different rounds. Higher naturalness signifies that the language model deems the sequence to be likely, w.r.t it's training data. We can see that as number of 3x binders increases going from round 2 to 3 or 4 the naturalness generally decreases. This signifies that as the designs are improving they are drifting slightly away from a naive antibody distribution.

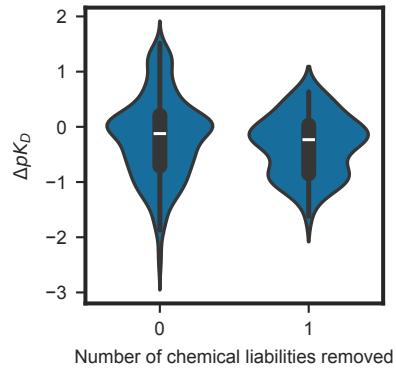


Fig. A8: Change in pK_D versus number of chemical liabilities removed from a seed sequence.

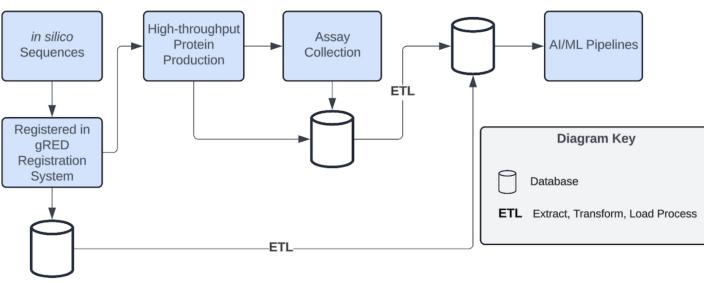
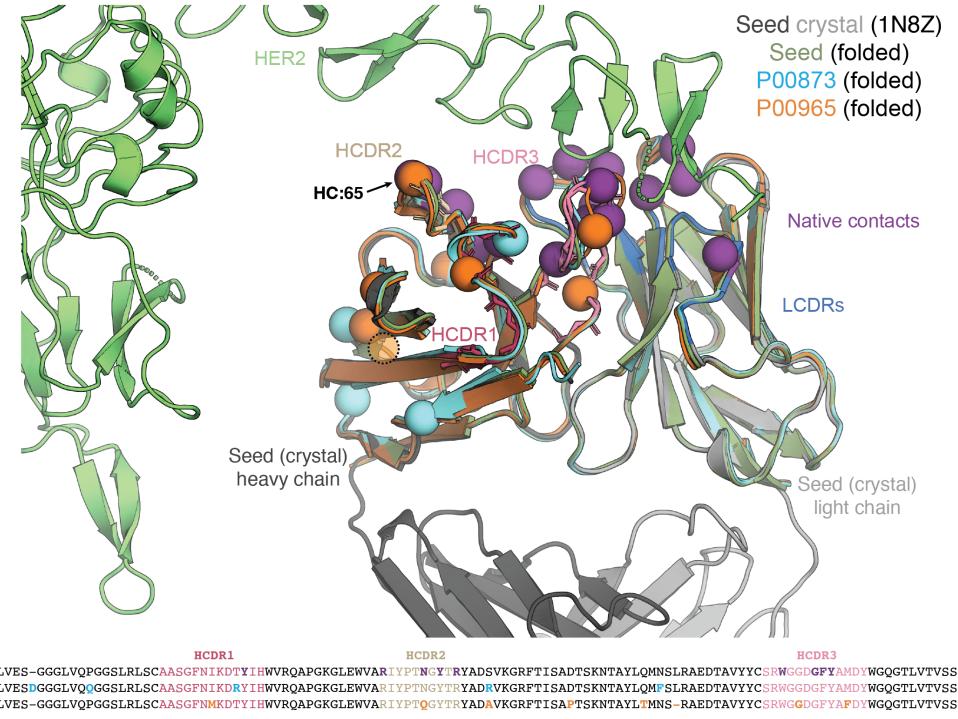


Fig. A9: Diagram of extract, transform, load process for data pipelines.



EVQLVES-GGGLVQPGGSLRLSC**AASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRADSVKGRFTISADTSKNTAYLQMNSLR**RAEDTAVYYC**SRWGGIGFYAMDW**WGQGTIVTVSS
EVQLVES**DGGGLVQGGSLRLSCAASGFNIKDRYIHWVRQAPGKGLEWVARIYPTNGYTRADSVKGRFTISADTSKNTAYLQMPSLR**AEDTAVYYC**SRWGGDGFYAMDW**WGQGTIVTVSS
EVQLVES-GGGLVQPGGSLRLSC**AASGFNMKD**TYIHWVRQAPGKGLEWVA**R**IYPT**QGYTR**YADA**V**KGRFTISAP**T**SKNTAYLT**M**NS-RAEDTAVYYC**SRWGGDGFYAFD**W**GQGTIVTVSS**

Fig. A10: Zoom-out of the Trastuzumab (hu4D5-linear)-HER2 complex crystal structure (PDBID: 1N8Z) with heavy/light chain colored in dark/light gray, HCDRs colored as in Fig 4, LCDRs in dark blue, native contacts in purple, with folded Fv structures (ABodyBuilder2) of the seed (dark green) and two designs (light blue and orange) aligned onto the crystal. A dashed transparent circle depicts the approximate location of the deletion in P00965. The folded structures exhibit very close similarities in all loops except for minor variability in HCDR3, indicating that the sequence changes are unlikely to have introduced dramatic structural rearrangements. Note that position 65 on the heavy chain (HC:65) is the only native seed contact that is directly mutated (in P00965 only). All other mutations are orthogonal to native contacts, with most localized to the framework region. The complete heavy-chain sequences for the seed and two designs are shown at bottom and color-coded by origin. There were no edits to the light chain in either of these two designs.

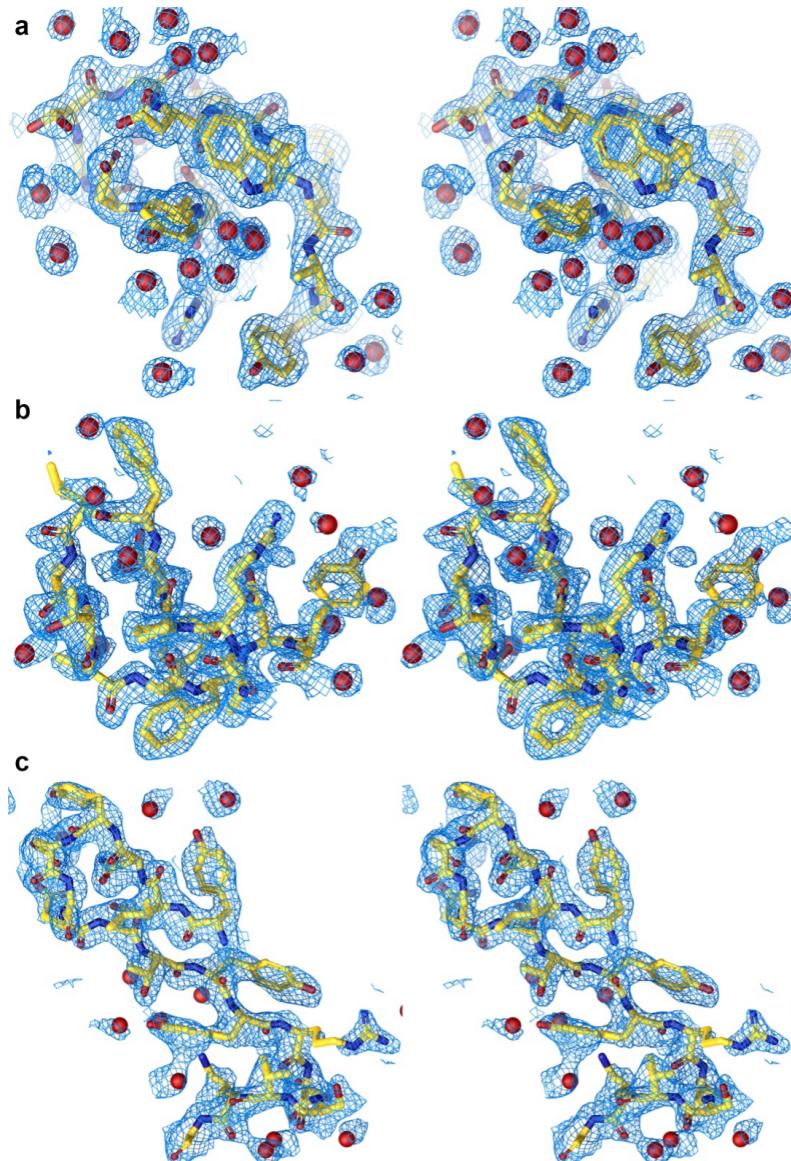
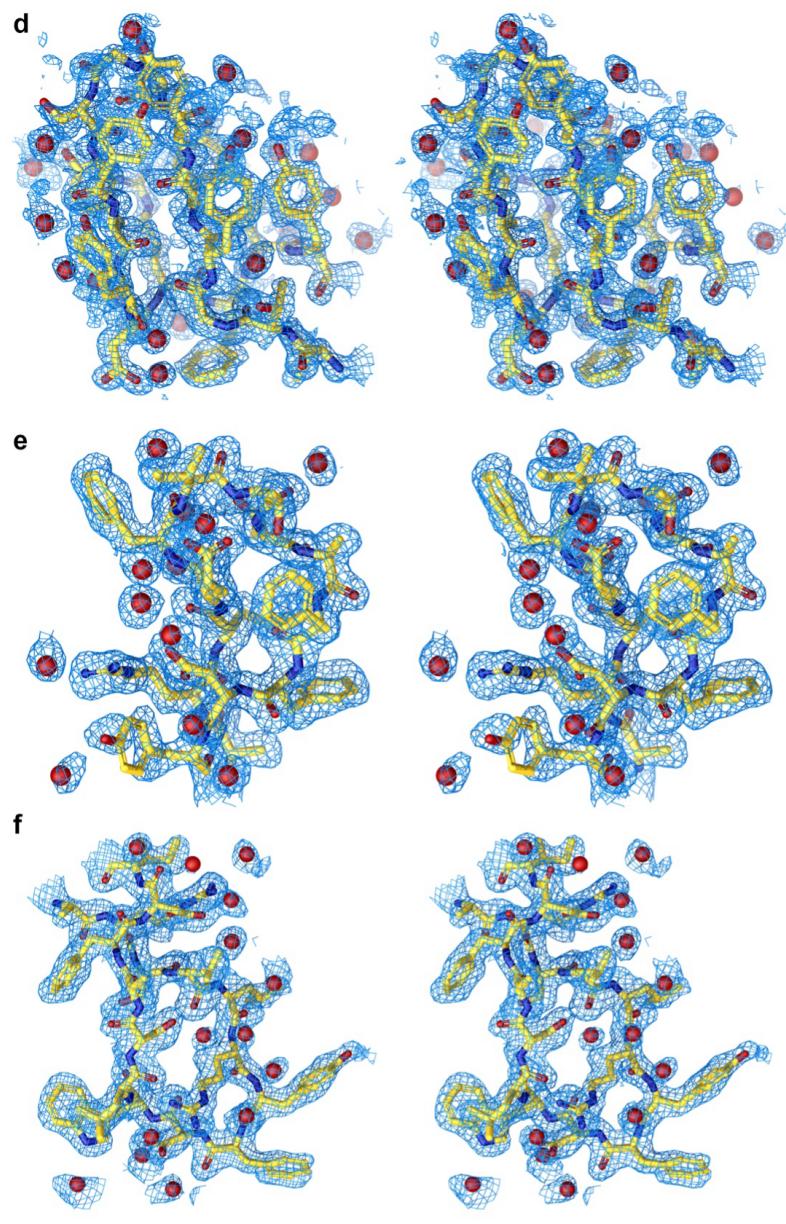
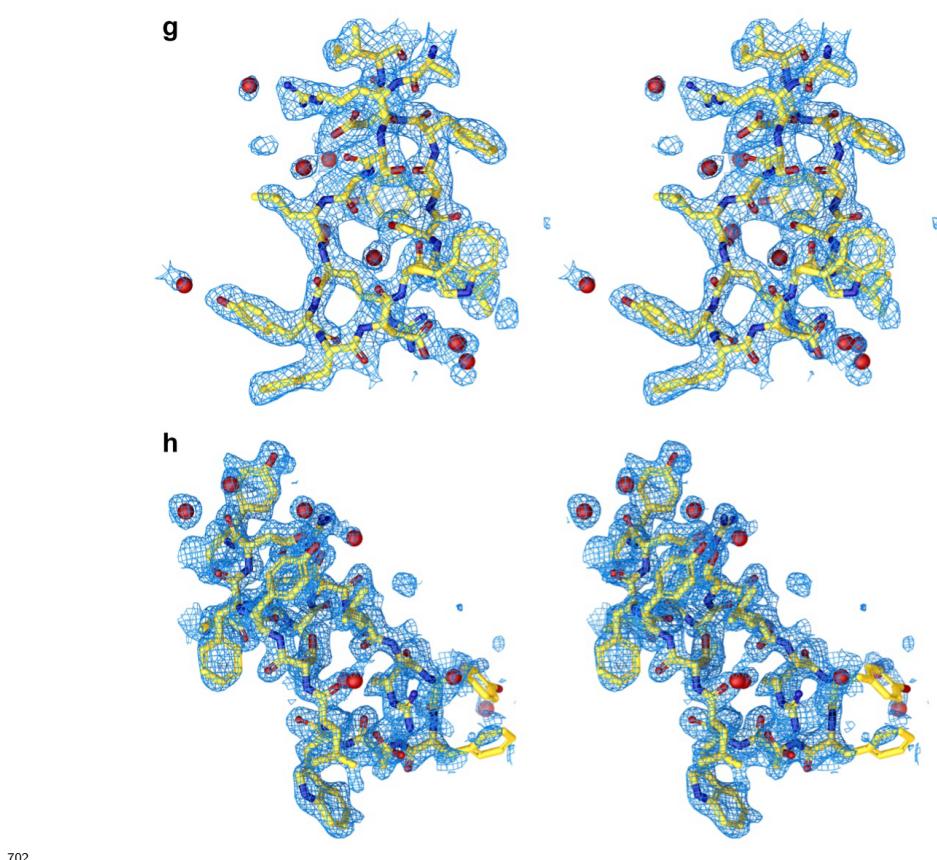


Fig. A11: Electron density maps. Divergent eye stereo images of the simulated annealing composite omit difference electron density maps, $(2m|F_o|-D|F_c|) \exp(iac)$, CDR H3 of Fabs **a** 9MVJ, **b** 9MUZ, **c** 9MUY (CDR H2), **d** 9MZC, **e** 9MZV, **f** 9N2V, **g** 9N7M, and **h** 9N7O, contoured at 1σ .



701



703 References

- 704 [1] Brekke, O.H., Sandlie, I.: Therapeutic antibodies for human diseases at the dawn
705 of the twenty-first century. *Nature reviews Drug discovery* **2**(1), 52–62 (2003)
- 706 [2] Carter, P.J., Rajpal, A.: Designing antibodies as therapeutics. *Cell* **185**(15),
707 2789–2805 (2022)
- 708 [3] Jain, S., Gupta, S., Patiyal, S., Raghava, G.P.: Thpdb2: compilation of fda
709 approved therapeutic peptides and proteins. *Drug Discovery Today*, 104047
710 (2024)
- 711 [4] Hsiao, Y.-C., Shang, Y., DiCara, D.M., Yee, A., Lai, J., Kim, S.H., Ellerman,
712 D., Corpuz, R., Chen, Y., Rajan, S., *et al.*: Immune repertoire mining for rapid
713 affinity optimization of mouse monoclonal antibodies. In: *MAbs*, vol. 11, pp.
714 735–746 (2019). Taylor & Francis
- 715 [5] McMahon, C., Baier, A.S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J.X.,
716 Erlandson, S.C., Hilger, D., Rasmussen, S.G., Ring, A.M., *et al.*: Yeast surface
717 display platform for rapid discovery of conformationally selective nanobodies.
718 *Nature structural & molecular biology* **25**(3), 289–296 (2018)
- 719 [6] Makowski, E.K., Kinnunen, P.C., Huang, J., Wu, L., Smith, M.D., Wang, T.,
720 Desai, A.A., Streu, C.N., Zhang, Y., Zupancic, J.M., *et al.*: Co-optimization of
721 therapeutic antibody affinity and specificity using machine learning models that
722 generalize to novel mutational space. *Nature communications* **13**(1), 3788 (2022)
- 723 [7] Carter, P.J., Quarmby, V.: Immunogenicity risk assessment and mitigation for
724 engineered antibody and protein therapeutics. *Nature Reviews Drug Discovery*,
725 1–16 (2024)
- 726 [8] Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V., Kayser, V.: The state-of-play
727 and future of antibody therapeutics. *Advanced drug delivery reviews* **122**, 2–19
728 (2017)
- 729 [9] Wu, Z., Johnston, K.E., Arnold, F.H., Yang, K.K.: Protein sequence design with
730 deep generative models. *Current opinion in chemical biology* **65**, 18–27 (2021)
- 731 [10] Olsen, T.H., Moal, I.H., Deane, C.M.: Ablang: an antibody language model for
732 completing antibody sequences. *Bioinformatics Advances* **2**(1), 046 (2022)
- 733 [11] Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z., Verkuil, R.,
734 Tran, V.Q., Deaton, J., Wiggert, M., *et al.*: Simulating 500 million years of
735 evolution with a language model. *bioRxiv*, 2024–07 (2024)
- 736 [12] Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M.,
737 Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., *et al.*: Large language models gen-
738 erate functional protein sequences across diverse families. *Nature Biotechnology*

- 739 **41**(8), 1099–1106 (2023)
- 740 [13] Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M.,
741 Zitnick, C.L., Ma, J., *et al.*: Biological structure and function emerge from scaling
742 unsupervised learning to 250 million protein sequences. *Proceedings of the*
743 *National Academy of Sciences* **118**(15), 2016239118 (2021)
- 744 [14] Ferruz, N., Höcker, B.: Controllable protein design with language models. *Nature*
745 *Machine Intelligence* **4**(6), 521–532 (2022)
- 746 [15] Jiang, K., Yan, Z., Di Bernardo, M., Sgrizzi, S.R., Villiger, L., Kayabolen, A.,
747 Kim, B., Carscadden, J.K., Hiraizumi, M., Nishimasu, H., *et al.*: Rapid in silico
748 directed evolution by a protein language model with evolvepro. *Science*, 6006
749 (2024)
- 750 [16] Hie, B.L., Shanker, V.R., Xu, D., Bruun, T.U., Weidenbacher, P.A., Tang, S., Wu,
751 W., Pak, J.E., Kim, P.S.: Efficient evolution of human antibodies from general
752 protein language models. *Nature Biotechnology* (2023)
- 753 [17] Li, L., Gupta, E., Spaeth, J., Shing, L., Jaimes, R., Engelhart, E., Lopez, R.,
754 Caceres, R.S., Bepler, T., Walsh, M.E.: Machine learning optimization of can-
755 didate antibody yields highly diverse sub-nanomolar affinity antibody libraries.
756 *Nature Communications* **14**(1), 3454 (2023)
- 757 [18] Angermueller, C., Mariet, Z., Jester, B., Engelhart, E., Emerson, R., Alipanahi,
758 B., Lin, C., Shikany, C., Guion, D., Nelson, J., *et al.*: High-throughput ml-guided
759 design of diverse single-domain antibodies against sars-cov-2. *bioRxiv*, 2023–12
760 (2023)
- 761 [19] Ruffolo, J.A., Gray, J.J., Sulam, J.: Deciphering antibody affinity matura-
762 tion with language models and weakly supervised learning. *arXiv preprint*
763 *arXiv:2112.07782* (2021)
- 764 [20] Ruffolo, J.A., Bhatnagar, A., Beazer, J., Nayfach, S., Russ, J., Hill, E., Hussain,
765 R., Gallagher, J., Madani, A.: Adapting protein language models for structure-
766 conditioned design. *bioRxiv*, 2024–08 (2024)
- 767 [21] Bennett, N.R., Coventry, B., Goreshnik, I., Huang, B., Allen, A., Vafeados, D.,
768 Peng, Y.P., Dauparas, J., Baek, M., Stewart, L., *et al.*: Improving de novo protein
769 binder design with deep learning. *Nature Communications* **14**(1), 2625 (2023)
- 770 [22] Bennett, N.R., Watson, J.L., Ragotte, R.J., Borst, A.J., See, D.L., Weidle, C.,
771 Biswas, R., Shrock, E.L., Leung, P.J., Huang, B., *et al.*: Atomically accurate de
772 novo design of single-domain antibodies. *bioRxiv*, 2024–03 (2024)
- 773 [23] Lisanza, S.L., Gershon, J.M., Tipps, S.W., Sims, J.N., Arnoldt, L., Hendel, S.J.,
774 Simma, M.K., Liu, G., Yase, M., Wu, H., *et al.*: Multistate and functional protein

- 775 design using rosettafold sequence space diffusion. *Nature biotechnology*, 1–11
776 (2024)
- 777 [24] Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M.R.,
778 Carotenuto, A., De Feo, G., Caponigro, F., Salomon, D.S.: Epidermal growth
779 factor receptor (egfr) signaling in cancer. *Gene* **366**(1), 2–16 (2006)
- 780 [25] Rossi, J.-F., Lu, Z.-Y., Jourdan, M., Klein, B.: Interleukin-6 as a therapeutic
781 target. *Clinical Cancer Research* **21**(6), 1248–1257 (2015)
- 782 [26] Yarden, Y.: Biology of her2 and its importance in breast cancer. *Oncology*
783 **61**(Suppl. 2), 1–13 (2001)
- 784 [27] Richards, C.D.: The enigmatic cytokine oncostatin m and roles in disease.
785 *International Scholarly Research Notices* **2013**(1), 512103 (2013)
- 786 [28] Ausserwöger, H., Schneider, M.M., Herling, T.W., Arosio, P., Invernizzi, G.,
787 Knowles, T.P., Lorenzen, N.: Non-specificity as the sticky problem in therapeutic
788 antibody development. *Nature Reviews Chemistry* **6**(12), 844–861 (2022)
- 789 [29] Notin, P., Rollins, N., Gal, Y., Sander, C., Marks, D.: Machine learning for
790 functional protein design. *Nature biotechnology* **42**(2), 216–228 (2024)
- 791 [30] Frey, N.C., Berenberg, D., Zadorozhny, K., Kleinhenz, J., Lafrance-Vanassee, J.,
792 Hotzel, I., Wu, Y., Ra, S., Bonneau, R., Cho, K., et al.: Protein discovery with
793 discrete walk-jump sampling. *arXiv preprint arXiv:2306.12360* (2023)
- 794 [31] Gruver, N., Stanton, S., Frey, N.C., Rudner, T.G., Hotzel, I., Lafrance-Vanassee,
795 J., Rajpal, A., Cho, K., Wilson, A.G.: Protein design with guided discrete
796 diffusion. *arXiv preprint arXiv:2305.20009* (2023)
- 797 [32] Lin, J.Y.-Y., Hofmann, J.L., Leaver-Fay, A., Liang, W.-C., Vasilaki, S.,
798 Lee, E., Pinheiro, P.O., Tagasovska, N., Kiefer, J.R., Wu, Y., Seeger,
799 F., Bonneau, R., Gligorijevic, V., Watkins, A., Cho, K., Frey, N.C.: Dyab: sequence-based antibody design and property prediction in a
800 low-data regime. *bioRxiv* (2025) <https://doi.org/10.1101/2025.01.28.635353>
<https://www.biorxiv.org/content/early/2025/02/02/2025.01.28.635353.full.pdf>
- 803 [33] Tagasovska, N., Gligorijević, V., Cho, K., Loukas, A.: Implicitly guided
804 design with propen: Match your data to follow the gradient. *arXiv preprint*
805 *arXiv:2405.18075* (2024)
- 806 [34] Daulton, S., Balandat, M., Bakshy, E.: Parallel Bayesian Optimization of Multiple
807 Noisy Objectives with Expected Hypervolume Improvement (2021)
- 808 [35] Wu, S., Tsukuda, J., Chiang, N., To, H., Chen, Y., Hötzl, I., Balasubramanian,
809 S., Nakamura, G., Kelly, R.L.: High titer expression of antibodies using linear

- 810 expression cassettes for early-stage functional screening. Protein Engineering,
811 Design and Selection, 012 (2024)
- 812 [36] Boder, E.T., Wittrup, K.D.: Yeast surface display for screening combinatorial
813 polypeptide libraries. Nature biotechnology **15**(6), 553–557 (1997)
- 814 [37] Raybould, M.I., Marks, C., Krawczyk, K., Taddese, B., Nowak, J., Lewis, A.P.,
815 Bujotzek, A., Shi, J., Deane, C.M.: Five computational developability guide-
816 lines for therapeutic antibody profiling. Proceedings of the National Academy of
817 Sciences **116**(10), 4025–4030 (2019)
- 818 [38] Hötzl, I., Theil, F.-P., Bernstein, L.J., Prabhu, S., Deng, R., Quintana, L., Lut-
819 man, J., Sibia, R., Chan, P., Bumbaca, D., *et al.*: A strategy for risk mitigation
820 of antibodies with fast clearance. In: MAbs, vol. 4, pp. 753–760 (2012). Taylor &
821 Francis
- 822 [39] Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak,
823 R., Kaufman, K.W., Renfrew, P.D., Smith, C.A., Sheffler, W., Davis, I.W.,
824 Cooper, S., Treuille, A., Mandell, D.J., Richter, F., Ban, Y.-E.A., Fleishman,
825 S.J., Corn, J.E., Kim, D.E., Lyskov, S., Berrondo, M., Mentzer, S., Popović,
826 Z., Havranek, J.J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray,
827 J.J., Kuhlman, B., Baker, D., Bradley, P.: Chapter nineteen - rosetta3: An
828 object-oriented software suite for the simulation and design of macromolecules.
829 In: Johnson, M.L., Brand, L. (eds.) Computer Methods, Part C. Methods in
830 Enzymology, vol. 487, pp. 545–574. Academic Press, ??? (2011). <https://doi.org/10.1016/B978-0-12-381270-4.00019-6> . <https://www.sciencedirect.com/science/article/pii/B9780123812704000196>
- 831
- 832
- 833 [40] Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park,
834 H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., Labonte, J.W.,
835 Pacella, M.S., Bonneau, R., Bradley, P., Dunbrack, R.L.J., Das, R., Baker, D.,
836 Kuhlman, B., Kortemme, T., Gray, J.J.: The rosetta all-atom energy function for
837 macromolecular modeling and design. Journal of Chemical Theory and Compu-
838 tation **13**(6), 3031–3048 (2017) <https://doi.org/10.1021/acs.jctc.7b00125> . PMID:
839 28430426
- 840 [41] Jeliazkov, J.R., Sljoka, A., Kuroda, D., Tsuchimura, N., Katoh, N., Tsumoto,
841 K., Gray, J.J.: Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity
842 Maturation Does Not Typically Result in Rigidification. Frontiers in Immunology
843 **9** (2018) <https://doi.org/10.3389/fimmu.2018.00413>
- 844 [42] Saremi, S., Hyvärinen, A.: Neural empirical Bayes. Journal of Machine Learning
845 Research **20**, 1–23 (2019)
- 846 [43] Kingma, D.P., Salimans, T., Poole, B., Ho, J.: Variational diffusion models.
847 In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in
848 Neural Information Processing Systems (2021)

- 849 [44] Honegger, A., PluÈckthun, A.: Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *Journal of molecular biology* **309**(3), 657–670 (2001)
- 850
851
- 852 [45] Rao, R.M., Liu, J., Verkuil, R., Meier, J., Canny, J., Abbeel, P., Sercu, T., Rives, A.: Msa transformer. In: International Conference on Machine Learning, pp. 8844–8856 (2021). PMLR
- 853
854
- 855 [46] Martinkus, K., Ludwiczak, J., LIANG, W.-C., Lafrance-Vanassee, J., Hotzel, I., Rajpal, A., Wu, Y., Cho, K., Bonneau, R., Gligorijevic, V., Loukas, A.: Abdifuser: full-atom generation of in-vitro functioning antibodies. In: Thirty-seventh Conference on Neural Information Processing Systems (2023). <https://openreview.net/forum?id=7GyYpomkEa>
- 856
857
858
859
- 860 [47] Liang, W.-C., Yin, J., Lupardus, P., Zhang, J., Loyet, K.M., Sudhamsu, J., Wu, Y.: Dramatic activation of an antibody by a single amino acid change in framework. *Scientific Reports* **11**(1), 22365 (2021)
- 861
862
- 863 [48] Fernández-Quintero, M.L., Kroell, K.B., Hofer, F., Riccabona, J.R., Liedl, K.R.: Mutation of framework residue h71 results in different antibody paratope states in solution. *Frontiers in Immunology* **12** (2021) <https://doi.org/10.3389/fimmu.2021.630034>
- 864
865
866
- 867 [49] Fernández-Quintero, M.L., Heiss, M.C., Liedl, K.R.: Antibody humanization—the Influence of the antibody framework on the CDR-H3 loop ensemble in solution. *Protein Engineering, Design and Selection* **32**(9), 411–422 (2020)
- 868
869
- 870 [50] Kelow, S., Faezov, B., Xu, Q., Parker, M., Adolf-Bryfogle, J., Dunbrack Jr, R.L.: A penultimate classification of canonical antibody cdr conformations. *bioRxiv*, 2022–10 (2022)
- 871
872
- 873 [51] Frey, N.C., Joren, T., Ismail, A., Goodman, A., Bonneau, R., Cho, K., Gligorijevic, V.: Cramming protein language model training in 24 gpu hours. *bioRxiv*, 2024–05 (2024)
- 874
875
- 876 [52] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., *et al.*: Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**(6637), 1123–1130 (2023)
- 877
878
- 879 [53] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M., Subramanian,
- 880
881
882
883
884
885
886

- 887 R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z.,
888 Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic,
889 R., Edunov, S., Scialom, T.: Llama 2: Open Foundation and Fine-Tuned Chat
890 Models (2023)
- 891 [54] Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H.: Uniref: com-
892 prehensive and non-redundant uniprot reference clusters. *Bioinformatics* **23**(10),
893 1282–1288 (2007)
- 894 [55] Olsen, T.H., Boyles, F., Deane, C.M.: Observed antibody space: A diverse
895 database of cleaned, annotated, and translated unpaired and paired antibody
896 sequences. *Protein Science* **31**(1), 141–146 (2022)
- 897 [56] Hauser, M., Steinegger, M., Söding, J.: Mmseqs software suite for fast and deep
898 clustering and searching of large protein sequence sets. *Bioinformatics* **32**(9),
899 1323–1330 (2016)
- 900 [57] Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Fried-
901 berg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.*: Biopython: freely
902 available python tools for computational molecular biology and bioinformatics.
903 *Bioinformatics* **25**(11), 1422–1423 (2009)
- 904 [58] Adebayo, J., Stanton, S.D., Kelow, S., Maser, M., Bonneau, R., Gligorijevic, V.,
905 Cho, K., Ra, S., Frey, N.C.: Identifying regularization schemes that make feature
906 attributions faithful. In: NeurIPS 2023 Workshop on New Frontiers of AI for
907 Drug Discovery and Development (2023)
- 908 [59] Tyka, M.D., Keedy, D.A., André, I., DiMaio, F., Song, Y., Richardson, D.C.,
909 Richardson, J.S., Baker, D.: Alternate states of proteins revealed by detailed
910 energy landscape mapping. *Journal of Molecular Biology* **405**(2), 607–618 (2011)
911 <https://doi.org/10.1016/j.jmb.2010.11.008>
- 912 [60] Maguire, J.B., Haddox, H.K., Strickland, D., Halabiya, S.F., Coventry, B., Griffin,
913 J.R., Pulavarti, S.V.S.R.K., Cummins, M., Thieker, D.F., Klavins, E., Szyper-
914 ski, T., DiMaio, F., Baker, D., Kuhlman, B.: Perturbing the energy landscape
915 for improved packing during computational protein design. *Proteins: Structure,*
916 *Function, and Bioinformatics* **89**(4), 436–449 (2021) <https://doi.org/10.1002/prot.26030> <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26030>
- 917 [61] Hsiao, Y.-C., Chen, Y.-J.J., Goldstein, L.D., Wu, J., Lin, Z., Schneider, K.,
918 Chaudhuri, S., Antony, A., Bajaj Pahuja, K., Modrusan, Z., *et al.*: Restricted epi-
919 tope specificity determined by variable region germline segment pairing in rodent
920 antibody repertoires. In: MAbs, vol. 12, p. 1722541 (2020). Taylor & Francis
- 921 [62] Shields, R.L., Namenuk, A.K., Hong, K., Meng, Y.G., Rae, J., Briggs, J., Xie,
922 D., Lai, J., Stadlen, A., Li, B., *et al.*: High resolution mapping of the binding site
923 on human IgG1 for FcγRI, FcγRII, FcγRIII, and FcRN and design of IgG1 variants with
924

- improved binding to the fc γ r. *Journal of Biological Chemistry* **276**(9), 6591–6604 (2001)
- [63] Kabsch, W.: xds. *Acta Crystallographica Section D: Biological Crystallography* **66**(2), 125–132 (2010)
- [64] McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J.: Phaser crystallographic software. *Journal of applied crystallography* **40**(4), 658–674 (2007)
- [65] Emsley, P., Cowtan, K.: Coot: model-building tools for molecular graphics. *Acta crystallographica section D: biological crystallography* **60**(12), 2126–2132 (2004)
- [66] Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczki, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al.: Macromolecular structure determination using x-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallographica Section D: Structural Biology* **75**(10), 861–877 (2019)
- [67] Maaten, L., Hinton, G.: Visualizing data using t-sne. *Journal of machine learning research* **9**(11) (2008)
- [68] Bachas, S., Rakocevic, G., Spencer, D., Sastry, A.V., Haile, R., Sutton, J.M., Kasun, G., Stachyra, A., Gutierrez, J.M., Yassine, E., et al.: Antibody optimization enabled by artificial intelligence predictions of binding affinity and naturalness. *bioRxiv* (2022)