

Framework choice preconfigures antibody developability

Chang Han Lee

chlee-antibody@snu.ac.kr

Seoul National University College of Medicine <https://orcid.org/0000-0002-7156-4423>

Woohyun Jeong

Seoul National University College of Medicine

Minho Park

Seoul National University College of Medicine <https://orcid.org/0009-0009-7496-6127>

Ji Hyun Choen

Seoul National University College of Medicine

Hyewon Lim

Seoul National University College of Medicine

Minjoo Kim

Seoul National University College of Medicine

Jisu Hong

Seoul National University College of Medicine

Youngju Kim

Seoul National University College of Medicine

Jisoo Yang

Seoul National University College of Medicine

Jimin Song

Seoul National University College of Medicine

Yujin Lee

Seoul National University College of Medicine

Yujin Jin

Seoul National University College of Medicine

Biological Sciences - Article

Keywords:

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11319929/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: **Yes** there is potential Competing Interest. C.-H.L., W.J. and M.P. are listed as inventors on Korean patent application no. 10-2025-0013450, which covers the automated antibody stability engineering and humanization software described in this work. The remaining authors declare no competing interests.

Framework choice preconfigures antibody developability

Woohyun Jeong^{1†}, Minho Park^{2†}, Ji Hyun Choen², Hyewon Lim², Minjoo MJ Kim^{2,3,4}, Jisu Hong^{2,3,4}, Youngju Kim^{2,3,4}, Jisoo Yang^{2,3,4}, Jimin Song^{2,3}, Yujin Lee^{2,3}, Yujin Jin^{2,3}, and Chang-Han Lee^{1,2,3,4,5,*}

¹ Cancer Research Institute, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

² Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

³ Department of Pharmacology, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

⁴ BK21 FOUR Biomedical Science Project, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

⁵ Convergence Dementia Research Center, Medical Research Center, Seoul National University, Seoul 03080, Republic of Korea

† These authors contributed equally to this work.

* Correspondence to Chang-Han Lee, Ph.D. Tel: +82-2-740-8278; email: chlee-antibody@snu.ac.kr

Antibody developability is still mostly repaired rather than designed^{1,2}. A binding domain is selected first, the framework carrying it is chosen for humanness^{3–5}, and its liabilities are corrected afterwards. Here we show that the framework decision itself carries this information. Antibodies that reach clinical development occupy a small, recurrent set of VH–VL germline pairings⁶. Within it, whether a transferred loop keeps its conformation is predicted by the local geometry at each complementarity-determining region (CDR)–

framework junction^{7,8}, the three framework and three CDR-terminal residues flanking it. CLade draws the recipient scaffold from the clinical set, ranks candidates by junction geometry rather than global similarity, and confines refinement to predefined positions. Using these criteria alone, we committed 23 designs across 16 antigen systems, spanning non-human donors and poorly developable human antibodies, to expression before any experimental measurement on them was available. Every design yielded recoverable protein and all 20 binding-evaluable designs stayed within tenfold of template EC₅₀. In addition, 21 of 23 met a multi-axis criterion defined before re-analysis, and 15 of 16 antigen systems qualified at the system level. Early developability engineering can therefore be moved upstream of expression, into the choice of framework.

Antibody developability liabilities include poor recombinant expression, conformational instability, aggregation, self-association and unfavourable solution behaviour^{9,10}. They typically become apparent only after a biologically active lead has been selected, when substantial discovery and functional-validation effort has already been spent^{1,2,10,11}. They are conventionally addressed by iterative mutagenesis, structural analysis and experimental screening aimed at repairing the original molecule while preserving its activity¹²⁻¹⁴. Computational design has greatly expanded that search. Inverse-folding models design sequences for a fixed backbone¹⁵, and repertoire-trained models optimize human-likeness within a given variable domain^{16,17}. Structure-generative models create new backbone architectures, which then require validation of manufacturability from scratch¹⁸. Each of these takes the variable-domain framework as given.

Recent de novo antibody design places newly designed CDR loops on a fixed, previously humanized framework, with broader developability properties such as aggregation, solubility and

expression remaining outside the primary design objective¹⁹. Similarly, a blinded, prospective benchmark of computational antibody discovery evaluated 511 designs from 29 organizations across three tasks, all of which varied CDRs within a given framework²⁰. The variable that decides whether a molecule survives development has not been the one under optimization.

A complementary strategy is to treat the framework itself as the primary design variable. Antigen recognition is largely encoded by the complementarity-determining regions (CDRs). The surrounding framework, however, governs variable-domain folding, VH–VL assembly, conformational stability, surface properties and presentation of the binding loops^{5,21–23}. Framework substitutions can tune stability and effector function through long-range structural coupling rather than acting as passive scaffolding²⁴. Framework replacement is long established^{25,26}, but recipient frameworks have historically been selected to reduce non-human sequence content, using germline identity, sequence similarity, structural compatibility, or repertoire prevalence^{3,4,17,27,28}, even though framework choice alone can determine which variants are recoverable²⁹. Where framework choice has been directed explicitly at developability, it has been guided by intrinsic physicochemical properties of germline frameworks or by empirical screening^{30,31}. We asked instead whether the frameworks of antibodies that have already advanced through therapeutic development constitute a usable prior. Clinical-stage and approved antibodies form a selected population whose members have passed molecular optimization, preclinical assessment, manufacturing development and, for clinical-stage molecules, human evaluation^{1,32,33}. Recurrence within this population does not establish the developability, immunogenicity or tolerability of a newly grafted molecule, which must be evaluated independently. It does define a therapeutic-use prior, a reweighting of the natural repertoire rather than a departure from it, concentrated on a small number of recurrent pairings^{34,35}.

Adopting such a prior shifts the central functional challenge to preservation of the parental paratope. CDR conformation is determined not only by the intrinsic sequence of each loop, but also by supporting framework interactions and by the local geometry through which the loop connects to the surrounding framework^{7,8,23,36}. For example, a single framework substitution at Vernier position H71 is sufficient to shift the conformational ensemble of several CDRs and the VH–VL orientation in solution³⁷. Perturbation of this boundary alters CDR backbone conformation, inter-CDR arrangement and side-chain presentation even when the grafted CDR sequence itself is unchanged^{8,27,38}. By contrast, global sequence identity and whole-variable-domain superposition are dominated by the substantially larger framework, and therefore underweight this local context^{33,39}.

Here we introduce CLade (Clinical-prior Local-Anchor-guided Developability Engineering). CLade uses V-gene representation and heavy–light pairing patterns from clinical-stage and approved antibodies to prioritize where to search for a recipient scaffold. It then ranks candidate scaffolds by the local geometry at each CDR–framework junction rather than by global framework similarity, and restricts subsequent sequence mutation to a predefined set of structurally supported positions. We term the resulting strategy developability grafting. The operation does not depend on the species of the donor. When the donor paratope is non-human, scaffold transfer reduces non-human sequence content as a consequence of selecting a clinically represented recipient, but the objective is the biophysical behaviour of the resulting molecule, not its humanness. The same operation applies unchanged to a human antibody whose own framework expresses poorly or aggregates. We evaluated both cases experimentally, generating 23 designs across 16 antigen systems assessed over 10 experimental assay classes (**Fig. 1a**).

In conventional workflows, these liabilities are identified only after expression, so each recovery cycle requires another round of expression (**Fig. 1b**). CLade runs as five computational operations, followed by experimental validation. These are a clinical prior database defining the recipient scaffold space, CDR-anchor alignment, structure-guided refinement, candidate validation and final sorting (**Extended Data Fig. 1**, with parameters in Methods). A full ranked candidate set is produced in 2 days.

A clinical prior compresses the space

CLade estimates its V–J germline pairing prior from a clinical reference set of antibodies that entered and advanced through clinical development. VH–VL germline pairs were assigned in the clinical reference set using IgBLAST⁴⁰ (**Fig. 2a**). The Observed Antibody Space (OAS) repertoire³⁴ enters only as an external natural-repertoire reference: OAS frequencies contribute neither to candidate inclusion nor to ranking.

VH–VL pair-frequency maps drawn on the same axes showed the clinical reference set concentrated on a small number of pairings that are rare in OAS (**Fig. 2b, c; Extended Data Fig. 2a, b**), and the three most frequent clinical VH–V κ pairs accounted for 15.89% of the clinical reference set but only 0.54% of OAS, whereas the three most frequent OAS pairs accounted for 4.29% of OAS and 1.98% of the clinical set.

The compression operates at the level of individual genes and of their pairings. The two sets share many frequently represented V genes but weight them very differently. The three most frequent IGHV genes in the clinical reference set accounted for 31.32% of heavy-chain assignments, compared with 15.84% in OAS, and for IGKV the corresponding totals were 55.63% and 35.24% (**Fig. 2d**). IGKV1-5 alone accounted for 38.4% of clinical light-chain assignments against 8.83% in OAS, and the single most represented IGHV and IGLV gene show the same asymmetry

(IGHV1-2 at 12.64% clinical versus 2.92% OAS, and IGLV2-14 at 22.58% clinical versus 14.27% OAS) (**Fig. 2d; Extended Data Fig. 2c**). To determine whether the pairing-level concentration exceeded that expected from the marginal VH and VL frequencies, we calculated a pseudocount-stabilized pairing-enrichment score, A_{ij} , relative to independent heavy–light pairing. The cumulative clinical-stage distributions were shifted toward positive enrichment compared with OAS for both kappa and lambda pairs (**Extended Data Fig. 2d, e**).

The number of unique VH–VL V-germline pairs observed in each reference set declines, from 3,270 in OAS to 128 in the most advanced clinical set (**Fig. 2e**). Expressed as proportions, pairs used by approved antibodies account for a progressively larger share at each stage (**Fig. 2f**). A protein language model embedding also places CLade sequences in a similar region of embedding space to the clinical reference set, rather than at its periphery (**Extended Data Fig. 2f**). Restricting the scaffold search to the clinical prior does not narrow the loops it must carry, and CDR-length distributions overlap for all six loops between OAS, the CLade prior and the clinical reference set (**Extended Data Fig. 3**).

Anchor geometry tracks loop conformation

We defined the anchor of each CDR as the three residues at each CDR terminus together with the three adjacent framework residues, six residues per junction and twelve per CDR (**Fig. 3a**). We refer to the six CDR-side residues as the CDR anchor, the six framework residues as the framework anchor, and the twelve together as the full anchor, and use only these three terms below. The hypothesis is geometric. If two structures present similar anchor geometry, the angles at which the loop leaves and re-enters the framework differ little and the loops they carry adopt similar

138 conformations, whereas divergent anchor geometry produces divergent loops even when the rest
139 of the framework is conserved.

140 To test this where the CDR is held constant by construction, we examined 173 CDR-loop
141 comparisons from 35 paired parental and humanized antibodies with unchanged CDR sequence,
142 binning each pair by the larger of its two absolute junction-angle deltas into four non-overlapping
143 bins defined as $\leq 1^\circ$, $>1^\circ$ to $\leq 2^\circ$, $>2^\circ$ to $\leq 3^\circ$, and $>3^\circ$. Median CDR root-mean-square deviation
144 (RMSD) increased monotonically across the four junction-angle-change bins, from 0.084 Å at 1
145 degree or less to 0.229 Å above 3 degrees (**Fig. 3b**).

146 Framework anchor identity was associated with preservation of these angles. Among structures
147 with identical CDRs, cases in which the six framework anchor residues were unchanged conserved
148 junction angles within 1 degree more than twice as often as cases in which they changed (37.2%
149 versus 17.9%). The advantage persists at the 2- and 3-degree thresholds, and is widest at 2 degrees
150 with a gap of 31.6 percentage points (**Fig. 3c**). The same relationship was observed across grafts
151 of the parental CDRs from the 35 humanization pairs from the clinical reference set onto distinct
152 recipient germ lines, with structures predicted using ColabFold^{41,42}, testing the compatibility of
153 each germline's intrinsic anchor context with the parental CDRs. Angle distortion increased with
154 framework-anchor substitution burden. Median angle delta rises from 2.117 to 4.210 degrees and
155 the fraction of junctions held within 3 degrees falls from 77.8% to 14.2% (**Fig. 3d**). Dataset-level
156 correlations were positive in 30 of 35 datasets (median $\rho = 0.644$; two-sided one-sample Wilcoxon
157 signed-rank test, $W = 34$, $n = 35$, $P = 2.17 \times 10^{-7}$). Integrated over all six CDRs, mean angle delta
158 and mean CDR RMSD covary across the full candidate set, from 0.195 Å at 1 degree or less to
159 0.582 Å above 3 degrees (**Extended Data Fig. 4a**), and 33 of 35 dataset-level correlations were

positive (median $\rho = 0.741$; two-sided one-sample Wilcoxon signed-rank test, $W = 4$, $n = 35$, $P = 4.07 \times 10^{-10}$).

Anchor sequence conservation and junction-angle conservation were coupled. The proportion of exactly matched framework anchors falls, and the proportion at Hamming distance 1 or 2 rises, across increasing junction-angle-change bins (**Extended Data Fig. 4b**). When the framework anchor does change, however, its replacement is not arbitrary. The substituted anchors concentrate on a small number of recurrent motifs per CDR, each close in Hamming distance to the template (**Extended Data Fig. 4c**). Scaffold selection under the clinical prior therefore tends to preserve template anchor character even where exact identity is lost, though unevenly across CDRs (**Extended Data Fig. 4d**).

CDR anchors showed the reciprocal restriction. Each framework anchor accommodated multiple CDR anchors, whereas each CDR anchor fit fewer framework anchors (**Extended Data Fig. 4e**).

Given this asymmetric mapping, we next examined whether CDR-side anchor identity further constrained junction geometry. Among structures sharing both CDR type and framework anchor, adding the CDR-side anchor raises the fraction falling into a dominant 3-degree angle bin from 50.9% to 71.5% at the N-terminal junction and from 45.7% to 67.7% at the C-terminal junction (**Extended Data Fig. 4f**). These gains exceeded those expected from group subdivision alone, estimated by shuffling CDR-anchor labels, by 9.1 and 8.8 percentage points, respectively.

To evaluate the effect of the CDR-side anchor itself, we compared matched loops sharing the same framework anchor but differing in CDR-side anchor identity. Such cases were sparse in the clinical reference humanized pairs, so we analysed SAbDab pseudo-pairs instead⁴³. Pairs with identical CDR-side anchors showed higher CDR core sequence identity, reaching a mean HCDR3 identity of 91.6% (**Fig. 3e**). When evaluated by full-anchor identity, pairs with an identical full anchor

showed lower mean CDR backbone RMSD than pairs with a non-identical full anchor for HCDR3 (0.61 Å versus 2.49 Å) (**Fig. 3f**). The same direction held for every CDR region and in all 19,717 region-specific leave-one-structure-out recalculations. The anchor-angle, angle-preservation, angle-bin-purity and mapping-multiplicity patterns were also reproduced in the experimentally determined SAbDab dataset (**Extended Data Fig. 5**), supporting their generalization beyond the clinical reference humanization pairs and predicted candidate structures.

Anchor alignment outperforms global superposition

We then compared three schemes for ranking candidate scaffolds. Oracle alignment superposed all six CDRs directly, Anchor alignment superposed the six full anchors, and Global alignment superposed the common Kabat-numbered backbone⁴⁴. The oracle is an upper reference rather than an independent ground truth, since it uses the CDR coordinates the other two schemes are trying to predict. Against it, anchor-based superposition reduces mean absolute error by 60.3% at the top-1 cutoff (two-sided paired Wilcoxon tests with Holm correction, $P = 2.38 \times 10^{-5}$) and by 66.6% at top-10 ($P = 7.02 \times 10^{-6}$). Reductions at the remaining cutoffs range from 61.9% to 65.7% (**Fig. 4a; Extended Data Fig. 6a**). Anchor alignment was closer to the reference for 28 of 35 comparisons at top-1 (Holm-adjusted paired Wilcoxon $P = 3.86 \times 10^{-4}$) and for 785 of 900 at top-30 ($P = 9.61 \times 10^{-6}$). Every intermediate cutoff is also significant (**Fig. 4b**).

Rank correlation with the oracle, and recovery of its top-ranked candidates, show the same separation. Mean Spearman ρ for Anchor exceeds Global at every cutoff, although the difference is not significant at top-5 (**Fig. 4c**). Anchor recovers more of the oracle-defined top-k at every cutoff except at top-1 (**Fig. 4d**). The practical consequence is illustrated by a representative tislelizumab candidate set, from which anchor alignment recovered 9 of the 10 reference top

candidates against 2 for global alignment (**Fig. 4e**). The two strategies are not interchangeable when only a few candidates can be advanced experimentally.

We next evaluated optimal anchor length. Absolute error against the reference falls steeply to 3+3 and then flattens (**Extended Data Fig. 6b**), but local framework RMSD around the anchor is minimized at 3+3 and rises at both 4+4 and 5+5 (**Extended Data Fig. 6c**). Absolute error continues to fall at 4+4 and 5+5, but direct CDR sequence coverage rises with window width (already 56.9% at 3+3, and 71.5% and 82.5% at 4+4 and 5+5), so wider windows converge on the reference rather than predicting it. The 3+3 anchor is therefore the widest window that remains local, and is used throughout.

Bounded refinement restores expression

Grafting alone can leave residual incompatibilities that require substitution. To keep the substitution budget small and the paratope intact, CLade restricts substitution positions to Vernier-zone and VH–VL interface residues that are poorly conserved across germlines and restore template-specific contacts. Their side chains are resolved against AlphaFold2-derived chi-angle distributions⁴⁵, and candidates are filtered on CDR-focused predicted local distance difference test (pLDDT) patterns and on interface and hydrophobicity risk (**Extended Data Fig. 7a**).

We followed one system, an anti-CD5 antibody, through the full trajectory. The grafting and refinement clones described below were made while the pipeline was being developed and are not part of the evaluated panel. The anti-CD5 design that enters the panel is the developability-engineering clone, generated after the pipeline was fixed. The initial grafting clone showed 1.83 Å HCDR3 all-atom RMSD and 0.405 Å interface RMSD relative to template. A first refinement clone, carrying 2 substitutions relative to the grafting clone, reduced HCDR3 RMSD to 1.02 Å but left interface RMSD at 0.366 Å. A second refinement clone, carrying 8 substitutions relative to the

grafting clone, brought HCDR3 RMSD to 0.44 Å and interface RMSD to 0.234 Å (**Fig. 5a; Extended Data Fig. 7b**). Structural superposition shows HCDR3 side chains Arg99, Tyr101, Trp103, Tyr104 and Phe105 returning to template positions across these rounds (**Fig. 5b**). Paratope recovery, however, came at the cost of expression. The second refinement clone yielded 90 mg/L against 3,720 mg/L for template, a 41-fold loss, and it sat outside the safe region of the hydrophobicity map (**Fig. 5c, d**). We therefore applied a developability-engineering mode, which resulted in four residue differences from germline replacement, namely L3 V to Q, L13 L to A, L85 V to T and H76 I to T (**Fig. 5e**). The resulting developability-engineering clone restored expression to 3,540 mg/L while retaining antigen binding (EC_{50} , 1.196 nM) (**Fig. 5d, f**). Across the successive repair steps, CLade reduced the full combinatorial space of possible substitutions to the small set of substitutions incorporated into each tested clone (**Extended Data Fig. 7b**). Local confidence profiles converged on the template pattern. The HCDR3 pLDDT minimum at H97 differed from the template value of 86 by 7 units in the first refinement clone and by 1 unit in the second (**Extended Data Fig. 7c**). The criterion is agreement with the template profile rather than a high absolute value, because both structures carry prediction error and similar local error is taken to indicate similar local context. Because pLDDT reports model confidence rather than absolute accuracy, we used these profiles to trigger local inspection rather than as a stand-alone ranking endpoint. Side-chain refinement is not cosmetic. The chi-aware top-ten set shared no candidate with the set generated without side-chain scoring and only two with the set generated without chi-angle scoring (**Extended Data Fig. 7d**). Side-chain-aware evaluation therefore discriminates among candidates with broadly similar backbone geometry.

Benchmark across 16 antigen systems

251 We expressed 23 CLade candidates spanning 16 antigen systems in two arms. In the
252 developability-engineering arm an existing antibody was redesigned for biophysical properties (n
253 = 8), and in the developability grafting arm, a parental paratope was transferred to a clinically
254 conditioned scaffold (n = 15). Each candidate was compared against its own template and against
255 a control set of unrelated antibodies produced in the same format. These 23 designs were selected
256 by the computational criteria alone and committed to expression before any experimental
257 measurement on them was available.

258 Before committing candidates to expression we applied a two-round risk evaluation derived from
259 a clinical reference set. Round 1 places each candidate in the plane of buried VH–VL interfacial
260 hydrophobicity and exposed hydrophobicity. Twenty of 23 fell in the low-risk region of the 3×3
261 grid and 3 were flagged (FZD10 B CL2, IL4 CL1 and LOX1 CL1) (**Fig. 6a**). Round 2 asks whether
262 interchain packing offsets that surface risk. All 23, including the 3 flagged, fall within the clinical
263 reference distribution in the VH–VL interface plane and meet the base-filter interface thresholds,
264 and the flagged candidates were retained under the patch-high rescue rule (**Fig. 6b**; Methods;
265 **Supplementary Methods S7**). The filter returned all 23 candidates as expressible and all 23
266 yielded recoverable protein, although two later fell below the intermediate yield gate of the
267 composite criterion. All 23 predictions of expressibility were correct, but candidates the filter
268 excluded were not carried forward, so its ability to identify failures remains untested.

269 Grafted loop conformations met both predefined cutoffs, with median HCDR3 backbone RMSD
270 below 1.0 Å and side-chain RMSD below 1.5 Å (**Fig. 6c**).

271 Framework-anchor substitution burden tracked structural preservation. The within-antibody
272 correlation between total framework-anchor burden and total CDR backbone RMSD was positive
273 in 14 of the 15 antigen systems evaluable for this within-system analysis (mean $r = 0.480$; median

274 $r = 0.612$; two-sided exact sign test, $k = 14$ of 15 , $P = 9.77 \times 10^{-4}$) (**Fig. 6d**). Stratifying systems
275 by burden and, reciprocally, by RMSD gave the same ordering, with 14 of 15 systems shifting in
276 the expected direction in both analyses (**Extended Data Fig. 8a, b**). Framework anchor
277 substitution burden tracked binding retention. In the CD5 series, the developability-engineering
278 clone that retained binding carried fewer framework anchor substitutions than the first refinement
279 clone, which lost measurable binding (**Fig. 6e**). All 20 binding-evaluable candidates retained EC_{50}
280 values within 10-fold of the template, with low total framework-anchor substitution burdens
281 ranging from 0 to 9 substitutions (**Fig. 6f**; **Extended Data Fig. 8c**, **Extended Data Fig. 9**). LOX1
282 CL1 and TAG72 CL1/TAG72 CL2 were excluded because a directly comparable EC_{50} fold could
283 not be defined. Framework-anchor burden is therefore a scaffold-selection prior rather than a
284 binding predictor. Backbone and side-chain comparison, restricted refinement and confidence
285 filtering act downstream, and designs that retain binding despite anchor substitution are recovered
286 by those steps.

287 Developability was assessed on nine axes: yield, size-exclusion chromatography (SEC) monomer
288 content, onset temperature (T_{onset}), first and second melting-transition temperatures (T_{m1} and T_{m2}),
289 polydispersity index (PDI), affinity-capture self-interaction nanoparticle spectroscopy (AC-SINS),
290 high-throughput hydrophobic interaction (HIC) assay and polyreactivity (**Fig. 6g–h**, **Extended**
291 **Data Fig. 8d–j**). Yield increased from template to candidate in the developability-engineering arm,
292 with several templates below the intermediate cutoff rising above the good cutoff. SEC monomer
293 content exceeded 95% for all candidates in both arms. Thermal, colloidal and hydrophobic
294 measures remained within the cutoff ranges across these seven axes. Polyreactivity at 100 nM IgG
295 against four polyreactivity targets was generally comparable to or lower than that of the template

in both arms. Overall, the developability-engineering arm showed a larger upward shift than the developability grafting arm.

Composite developability was then scored against the criterion defined in Methods. Two outcomes do not depend on how that composite is defined. All 23 designs yielded recoverable protein, and all 20 binding-evaluable designs remained within tenfold of template potency. Beyond these, 15 of 15 developability grafting designs passed on the absolute thresholds alone, 6 of 8 developability-engineering designs passed, and 21 of 23 passed overall (**Fig. 6i**). The antigen-system level outcome, where a system qualifies when at least one of its designs passes, is 15 of 16 systems (95% CI 71.7–98.9%). The two remaining failures, FZD10 B CL1 and LOX1 CL1, fell below the mandatory intermediate yield gate. When the developability-engineering arm was also held to the absolute thresholds without relative-improvement rescue, the overall outcome was 19 of 23 candidates and 14 of 16 antigen systems (Supplementary Table 1). Thresholds and control calibration are provided in Supplementary Tables 2 and 3, respectively, and the per-design outcome matrix is shown in **Extended Data Fig. 10a**.

A limited head-to-head comparison was possible on the three antigen systems for which CUMAb-designed comparators²⁸ were produced and run through the same experimental panel. Scored on the composite criterion used in **Fig. 6i**, the CLade designs qualified in all three systems and the CUMAb designs in one. The CD40 comparator fell below the mandatory binding gate (412-fold change in apparent EC₅₀) and the MUC1 comparator below the mandatory yield gate (**Extended Data Fig. 10b–d**). Both comparator failures are at mandatory gates, so neither depends on the scoring rule. Under the absolute rule the CLade MUC1 design is also scored as a failure, leaving two of three for CLade and one of three for the comparators. Per-axis differences were system-specific rather than systematic. The two routes were comparable on most axes and separated on

yield for MUC1, where the CUMAb design reached 940 mg/L and the CLade design reached 1,645 mg/L. With three shared systems this comparison is illustrative and is not treated as a benchmark of the two methods. Humanness scores were equivalent across the clinical reference set, tested CLade candidates and CUMAb comparators (**Extended Data Fig. 10e**). In silico, CDR3 RMSD did not separate CLade and CUMAb at the candidate level, each achieving the lowest value on a different subset of candidates (Supplementary Table 6). At the method level, mean all-atom RMSD was lower for CLade in HCDR3 even after excluding CD5, whereas differences in the other five CDR categories were modest (Table 1). Neither view is therefore treated as a ranking of the two methods.

Discussion

Scaffold choice is an early developability decision, not merely a humanness decision. Once a framework is fixed, much of what follows is repair, and every cycle of repair costs another expression. CLade moves that decision to the front. The recipient is drawn from scaffolds represented among clinical-stage and approved antibodies, ranked by the local geometry that predicts paratope compatibility, and refinement is confined to a predefined set of structurally supported positions. Selection takes 2 days (**Fig. 1b**). Repair does not disappear. What changes is where the decision sits, not how much modification it ultimately takes.

CLade was designed so that its two constraints address different risks. The clinical prior compresses the search space along a therapeutic-precedent axis, and anchor alignment orders the compressed space along a paratope-compatibility axis. The two constraints are motivated separately and were applied together. Their individual contributions to the experimental outcomes were not separated here. The residual that refinement then addresses is small. In the anti-CD5

342 series, restricted refinement restored paratope geometry, and subsequent developability
343 engineering restored expression to 95% of the template level through four substitutions. These
344 substitutions were nominated by a computational screen rather than by iterative mutagenesis.
345 The data are consistent with loop conformation being strongly constrained by the geometry of the
346 two junctions, with residues beyond them contributing only indirectly. The substitution burden
347 data make the same point from the other side, angle distortion rising with the number of
348 framework-anchor substitutions.

349 The three components of CLade act on different properties and should not be read as
350 interchangeable. The clinical prior conditions the recipient scaffold pool on therapeutic precedent.
351 Anchor geometry orders that pool by paratope compatibility, and therefore speaks to binding
352 retention rather than to expression or colloidal behaviour. Exposed hydrophobicity and interface
353 packing, evaluated separately and before expression, are what the pipeline uses to anticipate
354 expression risk, and restricted refinement acts on both. Structural agreement with a template is
355 accordingly a necessary condition for a scaffold-transfer design and not a sufficient one. Inverse-
356 folding models such as ProteinMPNN address a related but distinct design question, the
357 compatibility of a sequence with a specified backbone¹⁵. Recent work has used such scores to
358 predict productivity within the mutational landscape of a single antibody⁴⁶. CLade instead
359 addresses cross-scaffold selection, prioritizing recipient frameworks expected to preserve the
360 geometry of a transferred paratope. We therefore view inverse folding as complementary to, rather
361 than directly interchangeable with, scaffold selection.

362 The comparator analysis makes a point that is easy to miss. On *in silico* CDR RMSD, CLade and
363 CUMAb are not separable, each producing the lower value for a different subset of candidates

(Table 1, Supplementary Tables 4–6). *In silico* scoring is used here to reduce a large scaffold space to an experimentally tractable set. Developability itself is established experimentally. Several limitations bound these claims. The study evaluates the integrated CLade workflow. The contributions of the clinical prior and the anchor ranking were not independently resolved, since the two were applied together throughout. No matched conventional campaign was run either, so **Fig. 1b** describes typical published workflows rather than a head-to-head measurement. The anti-CD5 series was examined during method development and is reported as a development case study. The concentration of VH–VL pairings in the clinical reference set is used as a search prior and is not interpreted as evidence of an underlying pairing-selection mechanism. Several supporting structural analyses use predicted rather than experimental models, or rely on SAbDab structures with their own selection biases. The benchmark addresses early biophysical behaviour. Formulation, high-concentration behaviour, immunogenicity and pharmacokinetics were not assessed, and the effective sample size for generalization is 16 antigen systems. The selection-first logic is not specific to humanization. Bispecific format conversion and CAR binder domain exchange face the same trade-off between scaffold quality and functional transfer. Single-domain formats share the junction-geometry problem but not the VH–VL pairing prior. What CLade supplies is a general recipe. Build a prior from scaffolds that have already survived the relevant selection, and rank within that prior by the local geometry of the transferred CDR and its anchors, rather than by global similarity.

References

1. Jain, T. *et al.* Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 944–949 (2017).

- 387 2. Zhang, W. *et al.* Developability assessment at early-stage discovery to enable development
388 of antibody-derived therapeutics. *Antib. Ther.* **6**, 13–29 (2023).
- 389 3. Almagro, J. C. & Fransson, J. Humanization of antibodies. *Front. Biosci.* **13**, 1619–1633
390 (2008).
- 391 4. Dall’Acqua, W. F. *et al.* Antibody humanization by framework shuffling. *Methods* **36**, 43–
392 60 (2005).
- 393 5. Ewert, S., Honegger, A. & Plückthun, A. Stability improvement of antibodies for
394 extracellular and intracellular applications: CDR grafting to stable frameworks and
395 structure-based framework engineering. *Methods* **34**, 184–199 (2004).
- 396 6. Dudzic, P. *et al.* Conserved heavy/light contacts and germline preferences revealed by a
397 large-scale analysis of natively paired human antibody sequences and structural data.
398 *Commun. Biol.* **8**, 1110 (2025).
- 399 7. Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of
400 immunoglobulins. *J. Mol. Biol.* **196**, 901–917 (1987).
- 401 8. Foote, J. & Winter, G. Antibody framework residues affecting the conformation of the
402 hypervariable loops. *J. Mol. Biol.* **224**, 487–499 (1992).
- 403 9. Liu, J., Nguyen, M. D. H., Andya, J. D. & Shire, S. J. Reversible self-association increases
404 the viscosity of a concentrated monoclonal antibody in aqueous solution. *J. Pharm. Sci.* **94**,
405 1928–1940 (2005).
- 406 10. Xu, Y. *et al.* Structure, heterogeneity and developability assessment of therapeutic
407 antibodies. *mAbs* **11**, 239–264 (2019).

- 408 11. Xu, Y. *et al.* Addressing polyspecificity of antibodies selected from an in vitro yeast
409 presentation system: a FACS-based, high-throughput selection and analytical tool. *Protein*
410 *Eng. Des. Sel.* **26**, 663–670 (2013).
- 411 12. Geoghegan, J. C. *et al.* Mitigation of reversible self-association and viscosity in a human
412 IgG1 monoclonal antibody by rational, structure-guided Fv engineering. *mAbs* **8**, 941–950
413 (2016).
- 414 13. Bailly, M. *et al.* Predicting Antibody Developability Profiles Through Early Stage
415 Discovery Screening. *mAbs* **12**, 1743053 (2020).
- 416 14. Makowski, E. K. *et al.* Reduction of therapeutic antibody self-association using yeast-
417 display selections and machine learning. *mAbs* **14**, 2146629 (2022).
- 418 15. Dauparas, J. *et al.* Robust deep learning-based protein sequence design using
419 ProteinMPNN. *Science* **378**, 49–56 (2022).
- 420 16. Marks, C., Hummer, A. M., Chin, M. & Deane, C. M. Humanization of antibodies using a
421 machine learning approach on large-scale repertoire data. *Bioinformatics* **37**, 4041–4047
422 (2021).
- 423 17. Prihoda, D. *et al.* BioPhi: A platform for antibody design, humanization, and humanness
424 evaluation based on natural antibody repertoires and deep learning. *mAbs* **14**, 2020203
425 (2022).
- 426 18. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion.
427 *Nature* **620**, 1089–1100 (2023).
- 428 19. Bennett, N. R. *et al.* Atomically accurate de novo design of antibodies with RFdiffusion.
429 *Nature* **649**, 183–193 (2026).

- 430 20. Erasmus, M. F. *et al.* A blinded, prospective benchmark of in silico antibody discovery
431 anchored to experimental affinity and developability. *Nat. Biotechnol.*
432 <https://doi.org/10.1038/s41587-026-03238-6> (2026).
- 433 21. Masuda, K. *et al.* The role of interface framework residues in determining antibody
434 V(H)/V(L) interaction strength and antigen-binding affinity. *FEBS J.* **273**, 2184–2194
435 (2006).
- 436 22. Casaz, P. *et al.* Resolving self-association of a therapeutic antibody by formulation
437 optimization and molecular approaches. *mAbs* **6**, 1533–1539 (2014).
- 438 23. Fernández-Quintero, M. L., Heiss, M. C. & Liedl, K. R. Antibody humanization-the
439 Influence of the antibody framework on the CDR-H3 loop ensemble in solution. *Protein*
440 *Eng. Des. Sel.* **32**, 411–422 (2019).
- 441 24. Ng, J. C. F. *et al.* Tuning antibody stability and function by rational designs of framework
442 mutations. *mAbs* **17**, 2532117 (2025).
- 443 25. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the
444 complementarity-determining regions in a human antibody with those from a mouse.
445 *Nature* **321**, 522–525 (1986).
- 446 26. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for
447 therapy. *Nature* **332**, 323–327 (1988).
- 448 27. Bujotzek, A. *et al.* VH-VL orientation prediction for antibody humanization candidate
449 selection: A case study. *mAbs* **8**, 288–305 (2016).
- 450 28. Tennenhouse, A. *et al.* Computational optimization of antibody humanness and stability by
451 systematic energy-based ranking. *Nat. Biomed. Eng.* **8**, 30–44 (2024).

- 452 29. Wang, Y. *et al.* Comparison of ‘framework Shuffling’ and ‘CDR Grafting’ in humanization
453 of a PD-1 murine antibody. *Front. Immunol.* **15**, 1395854 (2024).
- 454 30. Spanke, V. A. *et al.* Balancing the extremes for antibody developability: hydrophobic and
455 electrostatic germline framework signatures for CDR-loop compensation. *mAbs* **18**,
456 2627669 (2026).
- 457 31. Gupta, P. *et al.* Matrixed CDR grafting: A neoclassical framework for antibody
458 humanization and developability. *J. Biol. Chem.* **300**, 105555 (2024).
- 459 32. Raybould, M. I. J. *et al.* Five computational developability guidelines for therapeutic
460 antibody profiling. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4025–4030 (2019).
- 461 33. Jarasch, A. *et al.* Developability Assessment During the Selection of Novel Therapeutic
462 Antibodies. *J. Pharm. Sci.* **104**, 1885–1898 (2015).
- 463 34. Olsen, T. H., Boyles, F. & Deane, C. M. Observed Antibody Space: A diverse database of
464 cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci.* **31**,
465 141–146 (2022).
- 466 35. Bashour, H. *et al.* Biophysical cartography of the native and human-engineered antibody
467 landscapes quantifies the plasticity of antibody developability. *Commun. Biol.* **7**, 922
468 (2024).
- 469 36. Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical
470 structures of immunoglobulins. *J. Mol. Biol.* **273**, 927–948 (1997).
- 471 37. Fernández-Quintero, M. L., Kroell, K. B., Hofer, F., Riccabona, J. R. & Liedl, K. R.
472 Mutation of Framework Residue H71 Results in Different Antibody Paratope States in
473 Solution. *Front. Immunol.* **12**, 630034 (2021).

38. Rhodes, E. R., Faris, J. G., Petersen, B. M. & Sprenger, K. G. Common framework mutations impact antibody interfacial dynamics and flexibility. *Front. Immunol.* **14**, 1120582 (2023).
39. Liu, C., Denzler, L. M., Hood, O. E. C. & Martin, A. C. R. Do antibody CDR loops change conformation upon binding? *mAbs* **16**, 2322533 (2024).
40. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).
41. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
42. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
43. Dunbar, J. *et al.* SAbDab: the structural antibody database. *Nucleic Acids Res.* **42**, D1140–1146 (2014).
44. Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132**, 211–250 (1970).
45. Cagiada, M., Thomasen, F. E., Ovchinnikov, S., Deane, C. M. & Lindorff-Larsen, K. AF2 χ : Predicting protein side-chain rotamer distributions with AlphaFold2. Preprint at <https://doi.org/10.1101/2025.04.16.649219> (2025).
46. Chun, C. *et al.* Landscape-scale navigation unlocks antibody CDR structural logic for AI-guided rescue and therapeutic optimization. Preprint at <https://doi.org/10.64898/2026.04.21.719857> (2026).

Tables

Table 1. System-level mean CDR all-atom RMSD for CLade and CUMAb designs across 15 antigen systems, excluding CD5.

CDR	Method	Heavy Chain RMSD (Å)	Light Chain RMSD (Å)
CDR1	CLade	0.529	0.702
	CUMAb	0.511	0.533
CDR2	CLade	0.823	0.653
	CUMAb	0.896	0.511
CDR3	CLade	0.772	0.680
	CUMAb	1.439	0.544

Figure legends

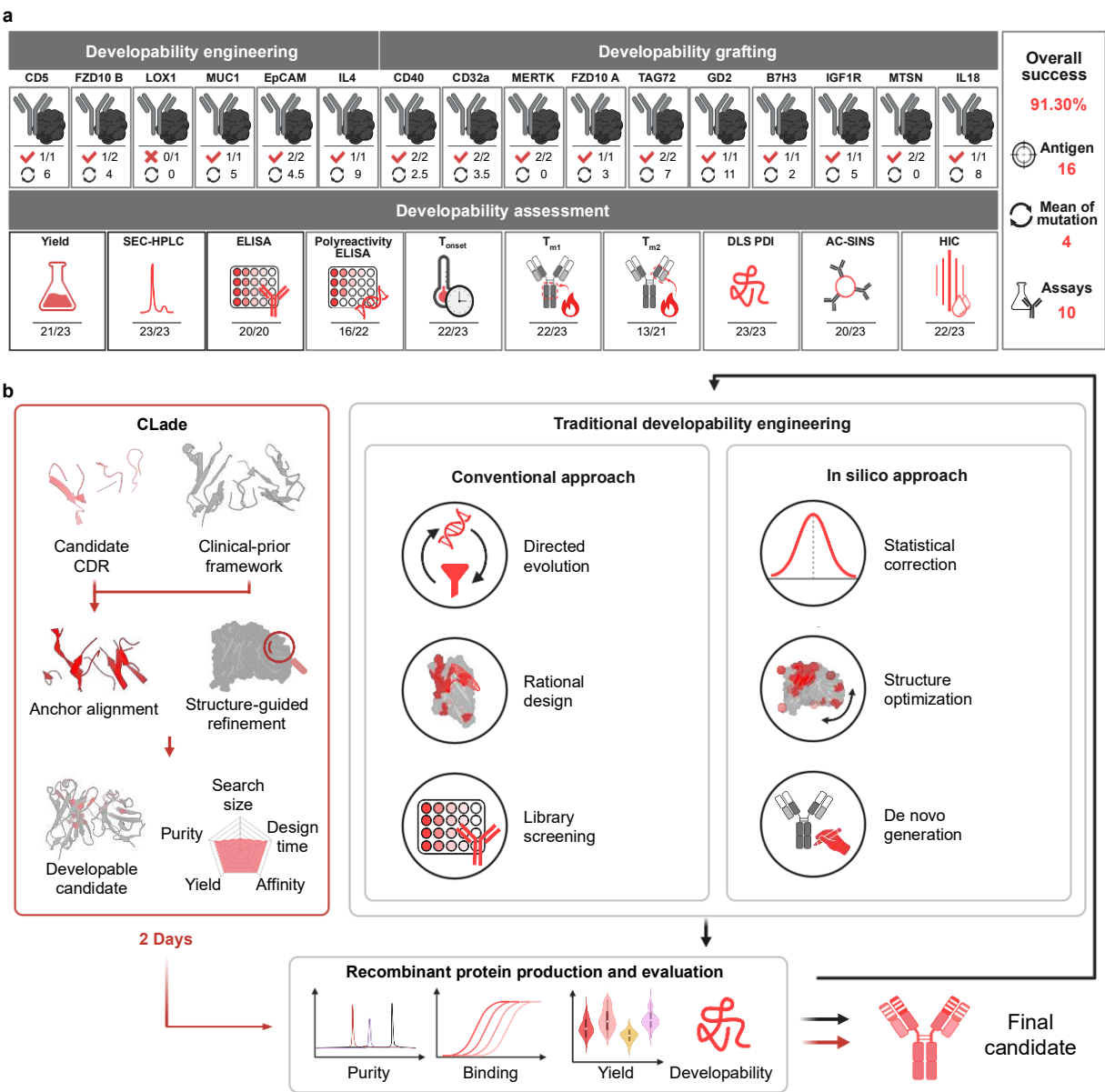


Figure 1. CLade moves scaffold selection and bounded refinement upstream of expression.

(a) Study-level summary of the design campaign across 16 antigen systems. The upper row separates developability-engineering and developability-grafting applications and reports the successful antigen-level designs and the mean number of introduced substitutions. The lower row lists the 10 assay classes used for developability assessment across 23 designs. The 91.30% value

shown in this schematic is the candidate-level composite outcome (21 of 23 designs) defined in Methods and reported in **Fig. 6i**. The corresponding system-level outcome, which counts a system as successful when at least one of its designs qualifies, is 15 of 16 systems. **(b)** Conceptual comparison of CLade with conventional experimental and in silico developability-engineering workflows. Conventional workflows begin with a fixed framework and repair liabilities through iterative design-build-test cycles. CLade instead transfers the candidate CDRs to a clinical-prior framework, evaluates local CDR-anchor compatibility, and performs restricted structure-guided refinement before expression.

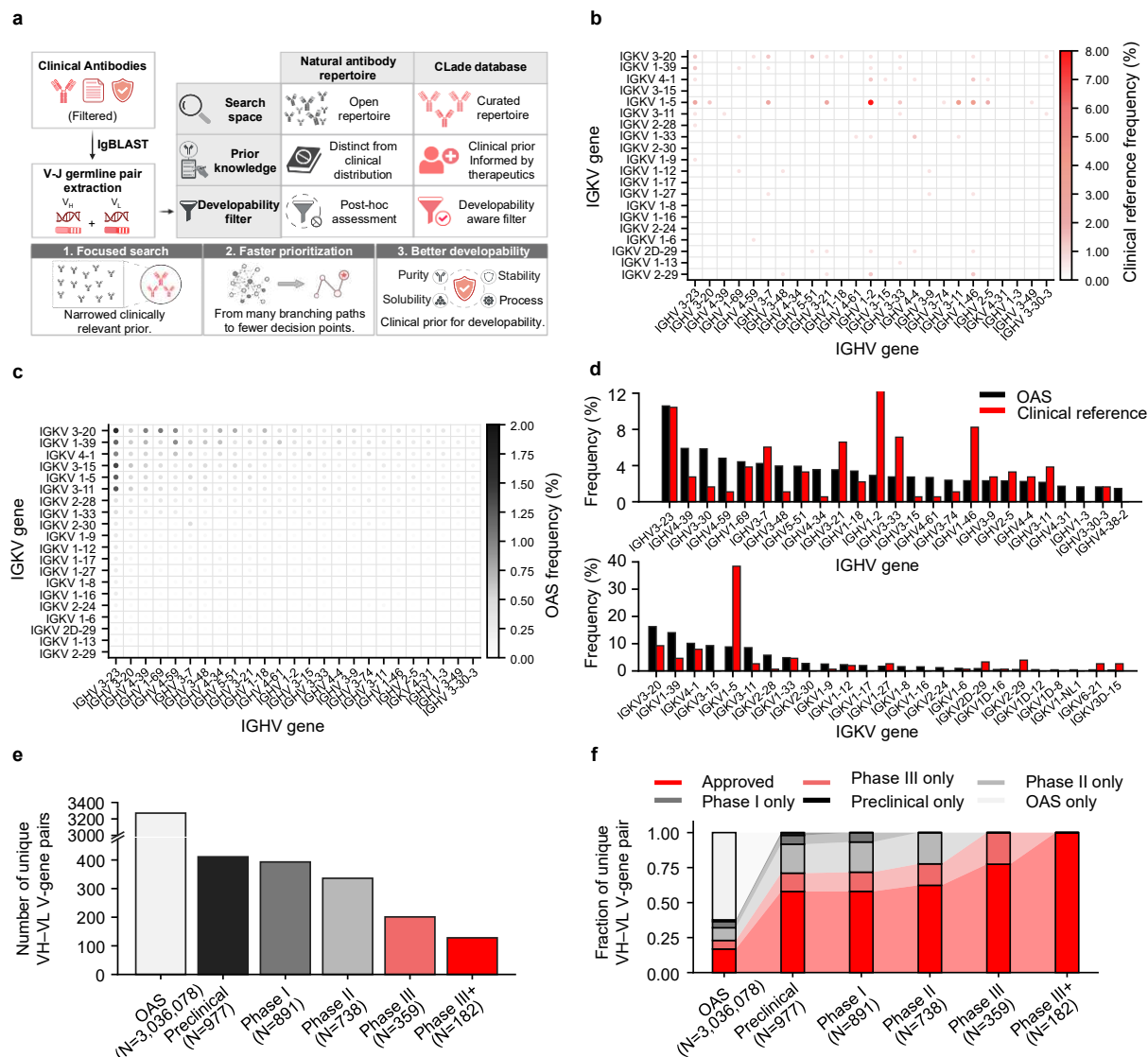


Figure 2. A clinical prior compresses the antibody V-germline pairing space.

(a) Construction and rationale of the CLade database. Clinically tested antibodies are annotated with IgBLAST, recurrent heavy- and light-chain V-germline pairs are extracted, and the resulting curated clinical repertoire is contrasted with broad OAS-based repertoire sampling. The lower schematics summarize focused search, faster prioritization, and developability-aware positioning.

(b) VH–V κ germline-pairing landscape in the clinical reference set. Bubble size and colour

525 intensity indicate the frequency of each heavy–light V-germline combination. **(c)** VH–Vκ
526 germline-pairing landscape in OAS, displayed on the same axes as in **(b)**. **(d)** IGHV (top) and
527 IGKV (bottom) gene-usage frequencies in OAS (black) and the clinical reference set (red). **(e)**
528 Number of unique V-germline pairs observed in OAS and in progressively advanced therapeutic-
529 development cohorts. **(f)** Stage-wise composition of V-germline pairs according to the most
530 advanced development category in which each pair is represented. Stacked fractions distinguish
531 approved, phase III-only, phase II-only, phase I-only, preclinical-only, and OAS-only pairs.
532

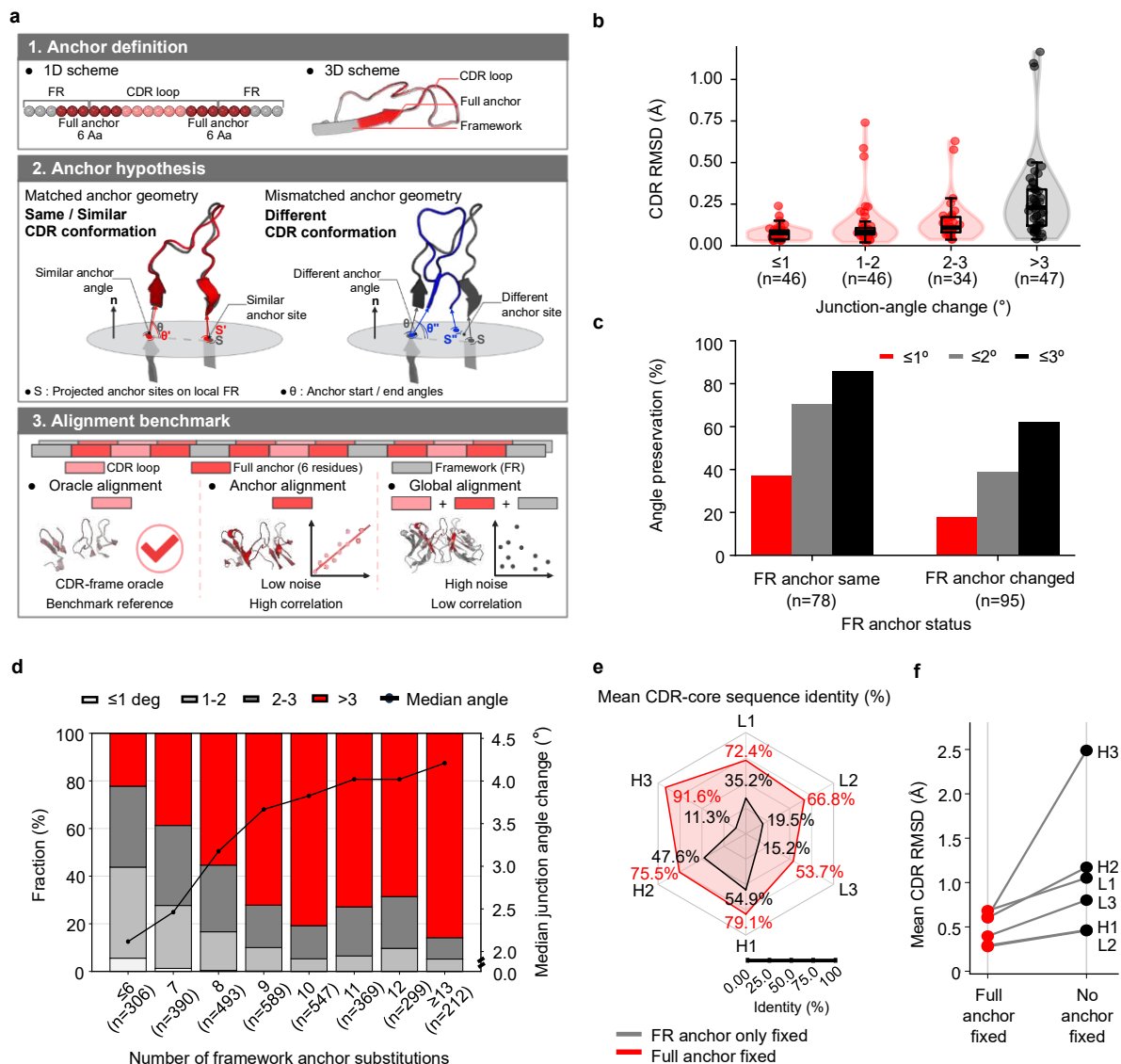


Figure 3. Full-anchor compatibility preserves CDR presentation.

(a) Anchor definitions and alignment benchmark. The framework anchor comprises three residues on each side of a CDR, the CDR anchor comprises three residues at each loop terminus, and the full anchor combines both. Oracle alignment uses all six CDRs, anchor alignment uses the six local full anchors, and global alignment uses the common Kabat-numbered variable-domain backbone.

(b) Full-CDR backbone RMSD for CDR-identical parental-humanized loop pairs stratified by the maximum N-terminal/C-terminal junction-angle change. Bins contain n = 46, 46, 34, and 47 loops

541 for ≤ 1 degree, 1–2 degrees, 2–3 degrees, and >3 degrees, respectively. **(c)** Junction-angle
 542 preservation among CDR-identical parental-humanized loop pairs with the framework anchor
 543 unchanged ($n = 78$) or changed ($n = 95$). Bars show the fractions with maximum N-terminal/C-
 544 terminal angle changes of ≤ 1 degree, ≤ 2 degrees, and ≤ 3 degrees. The corresponding preserved
 545 fractions are 37.2%, 70.5%, and 85.9% for unchanged anchors and 17.9%, 38.9%, and 62.1% for
 546 changed anchors. **(d)** Junction-angle composition and median angle change as a function of total
 547 framework-anchor substitution burden among 3,205 predicted candidates from 35 datasets.
 548 Stacked bars show ≤ 1 degree, 1–2 degrees, 2–3 degrees, and >3 degrees. The black line shows the
 549 median angle change. **(e)** Mean CDR-core sequence identity for each of the six CDRs in structure
 550 pairs with the framework anchor fixed alone (grey) or the full anchor fixed (red). **(f)** Mean CDR
 551 backbone RMSD for each loop in structure pairs with the full anchor fixed versus neither anchor
 552 component fixed. For H1, H2, H3, L1, L2 and L3, respectively, $n = 7,115, 17,075, 1,125, 49,487,$
 553 $56,461$ and $8,139$ with the full anchor fixed, versus $3,932,096, 2,616,019, 270,051, 822,863,$
 554 $3,803,848$ and $1,765,113$ with neither component fixed.

555

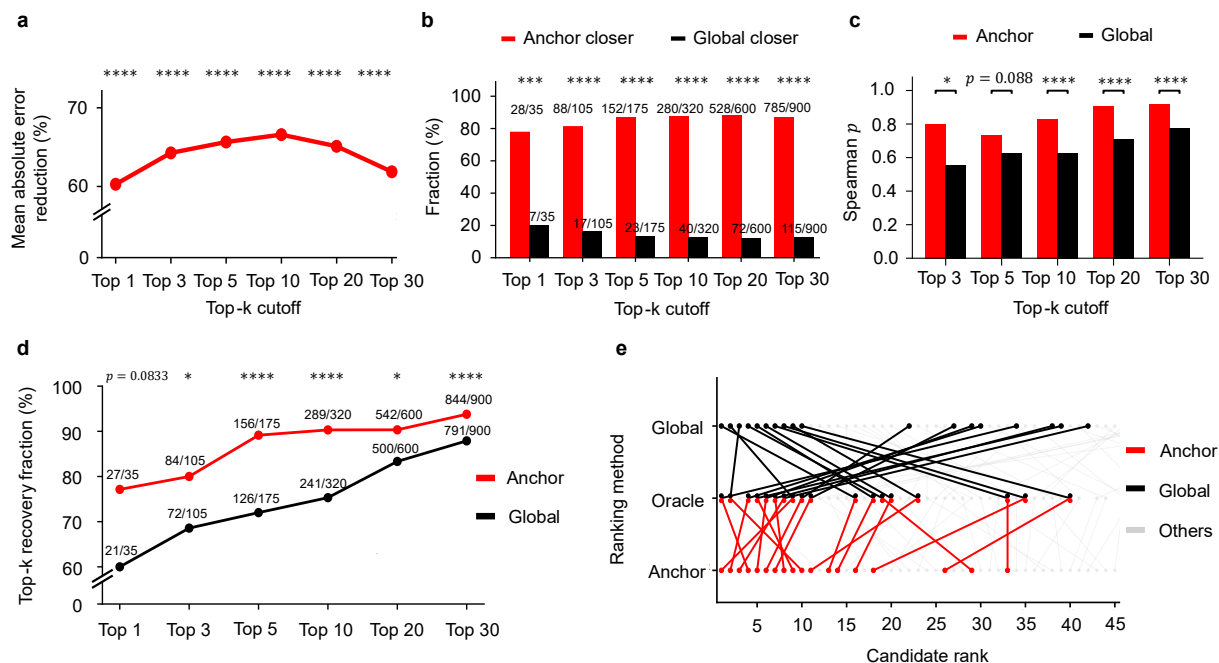


Figure 4. Full-anchor alignment improves recipient-scaffold ranking relative to global alignment.

(a) Reduction in mean absolute error from the oracle CDR-frame reference with anchor versus global alignment across top-k sets. Two-sided paired Wilcoxon signed-rank tests were Holm-corrected across cutoffs. For $k = 1, 3, 5, 10, 20$ and 30 , $n = 35, 35, 35, 32, 30$ and 30 ; $W = 57, 21, 10, 6, 2$ and 14 ; adjusted $P = 2.38 \times 10^{-5}, 7.02 \times 10^{-6}, 3.52 \times 10^{-6}, 7.02 \times 10^{-6}, 7.02 \times 10^{-6}$ and 1.40×10^{-5} , respectively. **(b)** Fraction closer to the oracle with anchor or global alignment; labels show pooled counts. For the same cutoffs and n , $W = 126, 47, 14.5, 9, 5$ and 4.5 ; adjusted $P = 3.86 \times 10^{-4}, 9.61 \times 10^{-6}, 2.39 \times 10^{-6}, 6.22 \times 10^{-6}, 8.60 \times 10^{-6}$ and 9.61×10^{-6} . **(c)** Mean within-dataset Spearman correlation with the oracle ranking for anchor and global alignment. For $k = 3, 5, 10, 20$ and 30 , $n = 35, 35, 32, 30$ and 30 ; $W = 3.5, 118, 72.5, 18$ and 14.5 ; adjusted $P = 0.0149, 0.0876, 0.00103, 4.10 \times 10^{-5}$ and 3.66×10^{-5} . **(d)** Oracle top-k recovery by anchor and global rankings, with pooled numerators and denominators. For $k = 1, 3, 5, 10, 20$ and 30 , $n = 35, 35, 35, 32, 30$ and 30 ; $W = 19.5, 34, 5, 12, 55.5$ and 6.5 ; adjusted $P = 0.0833, 0.0459, 3.93 \times 10^{-4}, 3.93 \times 10^{-4},$

571 0.0201 and 3.93×10^{-4} . **(e)** Ranking of 73 tislelizumab CLade candidates by global, oracle CDR
572 and anchor criteria. Anchor and oracle rankings share 9 of their top 10 candidates, compared with
573 2 of 10 for global and oracle rankings.

574

575

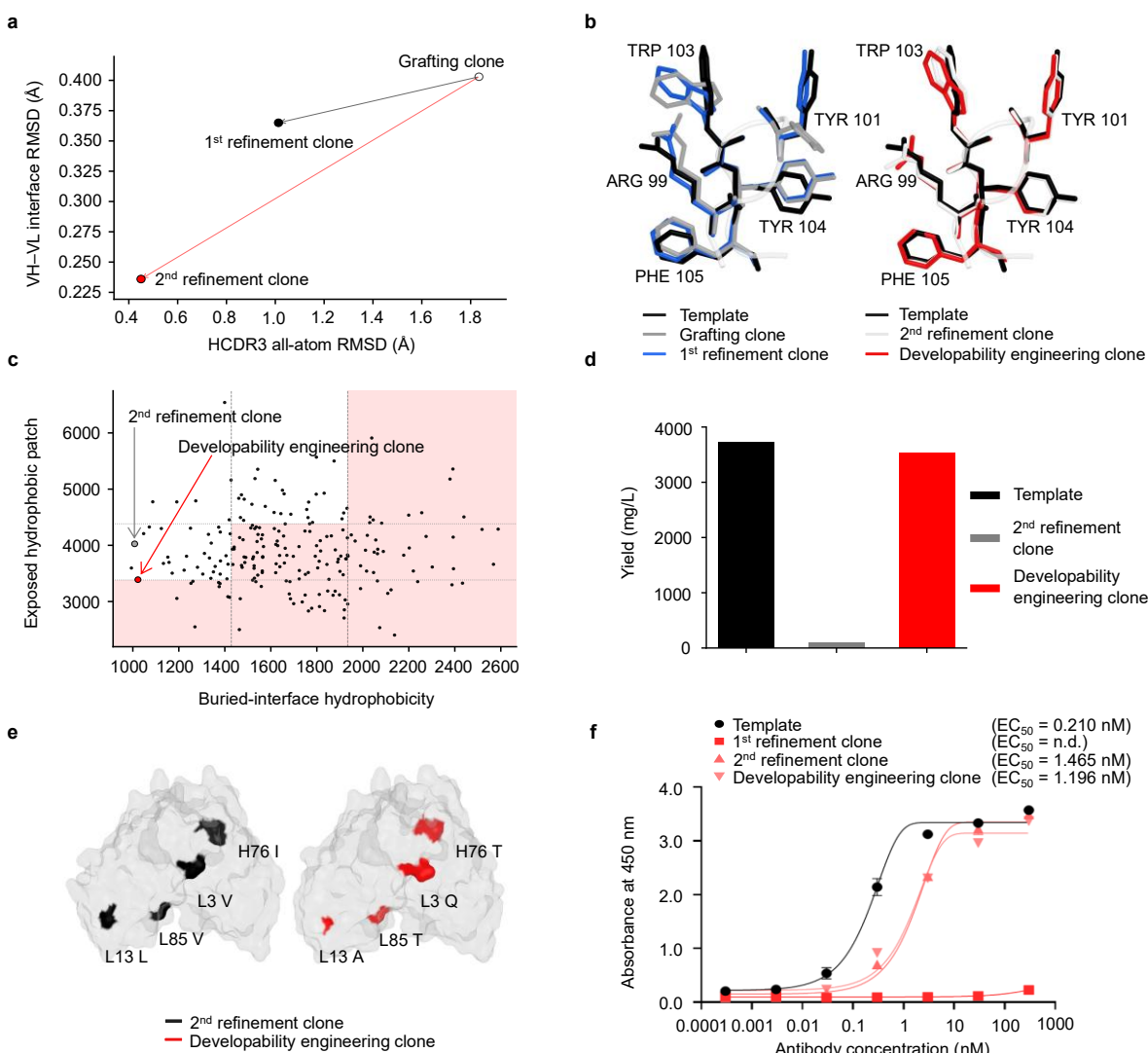


Figure 5. Restricted local refinement restores paratope geometry, binding, and expression in an anti-CD5 antibody.

(a) Structural trajectory of the anti-CD5 series in the plane of HCDR3 all-atom RMSD and VH-VL interface C α RMSD relative to the template. Points show the grafting clone, first-refinement clone, and second-refinement clone. Arrows indicate successive refinement steps. **(b)** Representative local overlays around HCDR3. The template is compared with the grafting and first-refinement clones (left) and with the second-refinement and developability-engineering

584 clones (right), showing progressive recovery of template-like side-chain packing. **(c)** Clinically
585 informed expression landscape defined by buried VH–VL interfacial hydrophobicity and exposed
586 non-interface hydrophobicity. The second-refinement and developability-engineering clones are
587 overlaid on the reference distribution. Dashed lines are the reference 20th and 80th percentiles and
588 pink regions denote the predefined low-risk zones. **(d)** Purified yield normalized to culture volume
589 was 3,720 mg/L for the template, 90 mg/L for the second-refinement clone, and 3,540 mg/L for
590 the developability-engineering clone. **(e)** Surface mapping of the four substitutions introduced
591 during developability engineering. The second-refinement clone (L3 V, L13 L, L85 V, and H76 I)
592 is compared with the developability-engineering clone (L3 Q, L13 A, L85 T, and H76 T). **(f)**
593 Antigen ELISA dose-response curves for the template, first-refinement, second-refinement, and
594 developability-engineering clones. Apparent EC₅₀ values are 0.210 nM, not determined, 1.465 nM,
595 and 1.196 nM, respectively.
596

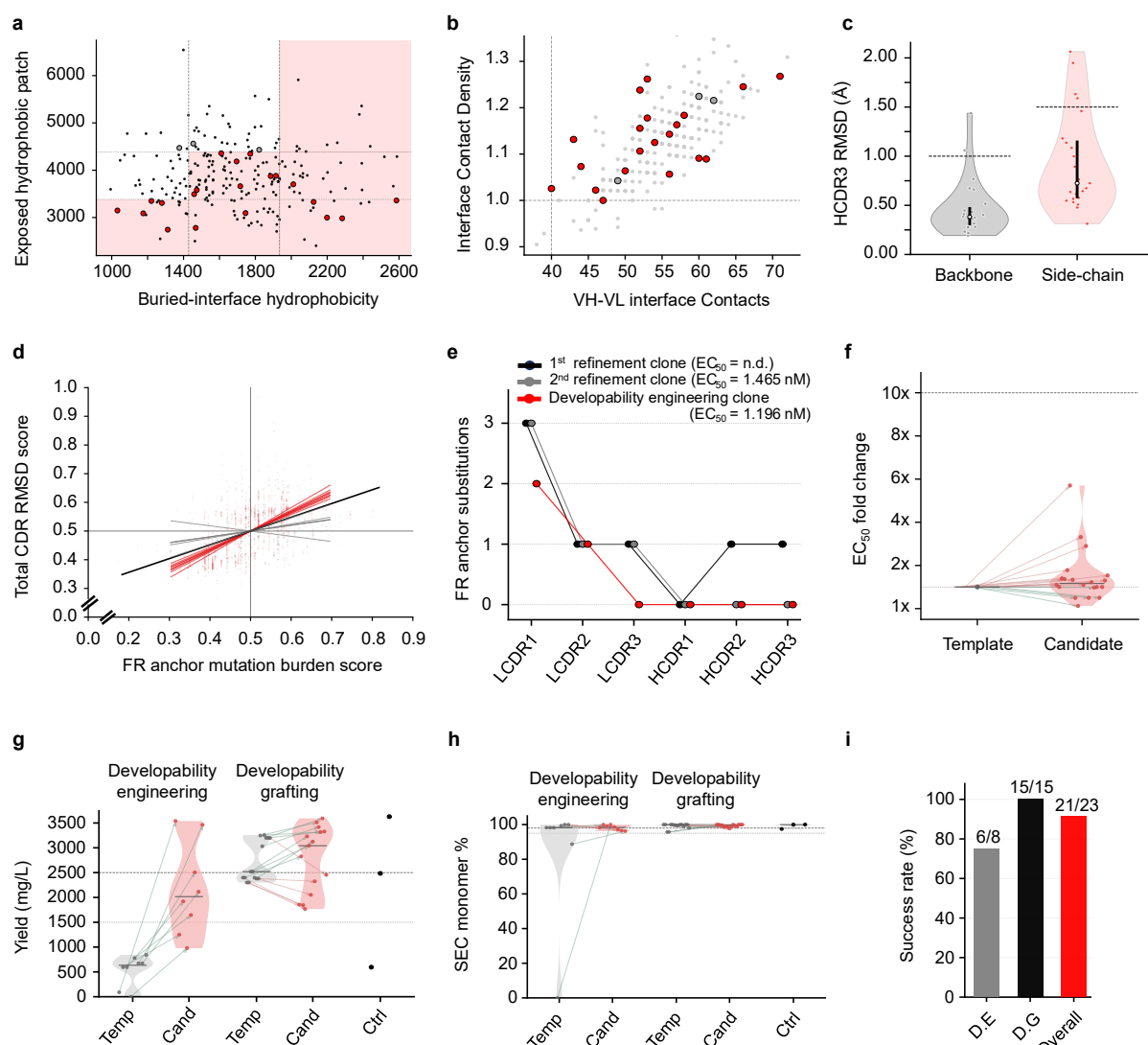


Figure 6. CLade preserves binding while supporting multi-axis developability across 23 evaluated candidates.

(a) Expression-risk landscape for 23 CLade candidates; 20 fall in predefined low-risk zones (red) and three are flagged (dark grey). (b) VH-VL interface packing versus clinical references (light grey), with low-risk candidates in red and flagged candidates in dark grey. Dashed thresholds are ≥ 40 contacts and ≥ 1.0 contact density; all 23 pass. (c) HCDR3 backbone and side-chain RMSD relative to template after anchor alignment ($n = 23$). Dashed lines mark 1.0 Å backbone and 1.5 Å

side-chain cutoffs. **(d)** Framework-anchor substitution burden versus CDR backbone RMSD in germline-only grafts, normalized within system. Lines show system-specific Pearson correlations; red indicates $r \geq$ the across-system mean (0.480) and grey indicates lower r . **(e)** Per-CDR framework-anchor substitution counts for the first-refinement, second-refinement, and developability-engineering anti-CD5 clones. Corresponding apparent EC_{50} values are not determined, 1.465 nM and 1.196 nM, respectively. **(f)** Apparent EC_{50} fold change relative to each cognate template for the 20 binding-evaluable candidates; for the anti-CD5 series the reference is the parental antibody (Methods). The dashed line marks the 10-fold acceptance criterion. All 20 evaluable candidates meet the criterion. LOX1 CL1 and TAG72 CL1/TAG72 CL2 are excluded because a directly comparable EC_{50} fold could not be defined. **(g,h)** Paired recombinant expression yield and size-exclusion chromatography (SEC) monomer content for template (Temp) and CLade candidate (Cand), shown separately for the developability-engineering and developability-grafting groups. Unrelated controls are shown in black; horizontal lines indicate predefined good and intermediate thresholds. **(i)** Composite developability success. Binding, yield and SEC monomer content are mandatory; up to two supporting failures are allowed, with polyreactivity counted once and missing values excluded. Success was 6/8 (75.0%), 15/15 (100%) and 21/23 (91.3%) for developability engineering, developability grafting and overall, respectively. Supporting-axis rescue applies only to the engineering arm; grafting designs are scored against absolute thresholds (Methods).

625 **Methods**

626 **Clinical prior construction.** The clinical reference set is the prior-construction cohort, the set of
627 molecules from which the CLade V-gene pairing prior is estimated: clinically tested antibodies
628 were collected from the Therapeutic Structural Antibody Database⁴⁷, with clinical-stage
629 annotations corresponding to the February 2025 database metadata, and filtered to human and
630 humanized molecules according to the predefined clinical-cohort genetics criteria. Heavy- and
631 light-chain sequences of the retained molecules were annotated with IgBLAST⁴⁰ and only
632 productive rearrangements were kept. The locus recorded for each light chain separated kappa
633 from lambda. V and J germline calls were matched against the IMGT reference directory to recover
634 the corresponding germline sequences⁴⁸, and heavy- and light-chain records were merged by
635 molecule name to give the paired V–J germline combination for each antibody, from which
636 separate kappa and lambda scaffold databases were built. The prior is estimated over V–J pairings.
637 The figures show V-gene pairings, because resolving J subdivides the space beyond legibility.
638 OAS sequences³⁴ were processed through the same IgBLAST-based germline-assignment
639 workflow rather than using germline calls supplied by OAS, and OAS was used only as an external
640 natural-repertoire reference; OAS-derived frequencies did not contribute to clinical-prior
641 construction, candidate inclusion or ranking. The broader stage-wise cohort spanned the clinical-
642 development stages shown in **Fig. 2e,f** and was not used to define the recipient-scaffold prior.

643 **Pairing-enrichment score.** For each VH–VL germline pair, the observed count Q_{ij} was
644 compared with the independent-pairing expectation $E_{ij} = N P(\text{VH}_i) P(\text{VL}_j)$. Here, N is the total
645 number of paired observations in the dataset, and $P(\text{VH}_i)$ and $P(\text{VL}_j)$ are the marginal
646 frequencies of the corresponding heavy- and light-chain germline genes in the same dataset. The

647 pseudocount-stabilized pairing-enrichment score was then calculated as $A_{ij} = \log_2 \left(\frac{Q_{ij}+0.5}{E_{ij}+0.5} \right)$.

648 Kappa and lambda pairs were analysed separately. The score describes concentration relative to
649 marginal usage. Cumulative clinical-stage distributions were compared with OAS using
650 empirical cumulative distribution functions (**Extended Data Fig. 2d, e**). Detailed stage pooling
651 and distributional procedures are provided in **Supplementary Methods S8**.

652 **CDR identification and grafting.** For each input antibody the heavy- and light-chain variable-
653 domain sequences were extracted, numbered with AbNumber⁴⁹ and partitioned into framework
654 and CDR segments using the Kabat definition⁴⁴. The six extracted CDRs were transferred onto the
655 framework regions of every scaffold in the clinical-prior database of the matching light-chain class,
656 generating the initial CDR-grafted candidate set. No residual substitutions were introduced at this
657 stage; all sequence modifications were confined to the refinement step described below.

658 **Humanness scoring.** Humanness of the grafted candidates and of the refinement variants
659 generated from them was scored as BioPhi OASis identity¹⁷ after Kabat numbering and Kabat
660 CDR assignment of both chains. Candidates at or above an OASis identity of 0.75 were
661 preferentially retained. If no candidate met this threshold, the cutoff was relaxed to 0.70.
662 Humanness was used as an inclusion filter only and does not enter the final ranking score.

663 **Structure prediction.** Paired Fv structures used throughout the CLade pipeline and downstream
664 structural analyses were generated locally with ColabFold⁴² and AlphaFold2⁴¹, except where
665 experimentally determined structures were used as indicated below. Heavy- and light-chain
666 variable domains were supplied as paired FASTA records and multiple-sequence alignments were
667 generated with MMseqs2^{50,51}. For pairwise structural comparisons, relaxed models from rank 1 to
668 a user-defined maximum rank were collected for each variant, all cross-rank combinations were

evaluated, and the model pair with the lowest mean RMSD across the six CDRs was retained as the representative structural pair (**Supplementary Methods S1**).

Anchor definition. The anchor of each CDR comprises the terminal CDR residues and the adjacent framework residues at each junction. To optimize the anchor-window length, symmetric windows ranging from 1+1 to 5+5 residues per junction were evaluated based on CDR coverage, CDR RMSD error and framework anchor RMSD. Based on the joint performance of these metrics, a 3+3 window was selected as the default anchor and used throughout subsequent CLade analyses. Details of the anchor-window optimization are provided in **Supplementary Methods S15**.

Junction angle and CDR RMSD. Angles were calculated independently in the original coordinate frame of each structure, whereas RMSD was calculated after optimal superposition. For each junction, the N-terminal angle was defined as the angle between the C α axis of the three framework residues immediately preceding the CDR and the C α axis of the first three CDR residues, with the C-terminal angle defined symmetrically. Differences in local junction geometry between template and candidate structures were quantified as the absolute differences in the N-terminal and C-terminal angles. CDR RMSD was calculated after optimal superposition of corresponding backbone atoms within each CDR. Each three-residue block was represented by a unit vector from the first to the third C α atom, such that the angle depended on segment direction rather than segment length. Full details of the calculations are provided in **Supplementary Methods S9**.

Angle bin purity. Absolute junction angles were binned in 3-degree increments. Framework-anchor-only purity was computed as the fraction of structures within a framework anchor group falling in the modal bin. Full-anchor purity was computed by further partitioning each framework anchor group by CDR-side anchor sequence and taking the subgroup-size-weighted mean of

within-subgroup modal fractions. CDR-anchor subgroups containing a single observation were excluded, as were framework-anchor groups with fewer than two remaining subgroups. CDR-anchor labels were shuffled 10,000 times within each retained framework-anchor group, preserving subgroup sizes, to estimate the purity increase expected from group subdivision alone. Detailed definitions and statistical procedures are provided in **Supplementary Methods S10**.

Reference structure cohort and CDR-level pseudo-pairing. The experimental structure cohort used in **Fig. 3e, f** and **Extended Data Fig. 5** was assembled from SAbDab⁴³. Structures were retained when both heavy and light chains were annotated and the resolution was 3.0 Å or better, and were then standardized to a common numbering and chain convention. Records with identical VH and VL sequences were collapsed to a single representative, and contained or truncated records were merged into the same representative group. CDR-level pseudo-pairs were then formed between distinct representative records sharing the same CDR region and length, and these pseudo-pairs were the unit of the anchor-sequence, boundary-geometry and CDR-preservation analyses. Anchor and CDR sequence identity, Hamming distance and anchor-excluded CDR-core definitions are provided in **Supplementary Methods S11**.

Leave-one-structure-out robustness analysis. For each CDR region, each unique structure contributing to the full-anchor-fixed or neither-anchor-fixed pseudo-pair group was removed in turn, and all pseudo-pairs containing that structure were excluded. Mean CDR backbone RMSD was then recalculated for both groups, and the between-group difference was expressed such that positive values indicated lower RMSD in the full-anchor-fixed group. The sign of this difference was recorded for each leave-one-structure-out recalculation, and the procedure was applied independently to H1, H2, H3, L1, L2 and L3.

Alignment schemes and ranking benchmark. Oracle alignment superposed all six CDRs and served as the CDR-frame reference. Anchor alignment superposed the six CLade anchors. Global alignment superposed the common Kabat-numbered backbone. Anchor and Global alignment were evaluated on the same candidate structures within each antibody dataset against the oracle reference. Alignment error and top-k recovery were summarized at each evaluable top-k cutoff ($k = 1, 3, 5, 10, 20$ and 30). Rank preservation was assessed by Spearman correlation where defined. All comparisons used the same dataset-level paired framework. The tests are specified under Statistics and the unit of analysis. Detailed metric definitions and statistical procedures are provided in **Supplementary Methods S12–S14 and S16**.

Refinement. Candidate substitution positions required germline conservation below 80% at the aligned position across the top 5 germlines of the corresponding V gene (IGHV for heavy-chain positions, IGKV or IGLV for light-chain positions), together with a template-specific contact. They were further restricted to Vernier-zone residues (heavy: 2, 27, 28, 29, 30, 47, 48, 49, 67, 69, 71, 73, 78, 93, 94, 103; light: 2, 4, 35, 36, 46, 47, 48, 49, 64, 66, 68, 69, 71, 98)⁸ and to VH–VL interface residues within 4 Å. The five recipient scaffolds with the lowest anchor-based alignment scores were considered for refinement. AF2 χ -supported χ_1 and χ_2 states⁴⁵ were used to identify side-chain conformations requiring refinement. Single and combinatorial substitutions at the retained positions generated the refinement variants, after which structural alignment and CDR RMSD calculations were repeated. Detailed interface-difference classification and side-chain refinement procedures are provided in **Supplementary Methods S2 and S3**.

Confidence filtering. Residue-level pLDDT values were taken from the ColabFold score output and CDRs were delimited by Kabat numbering. Within the model ensemble of one candidate, models whose mean pLDDT for a given CDR fell 6 or more points below the ensemble median

for that CDR were discarded as low-confidence outliers. The median pLDDT at each Kabat-matched CDR position across the surviving models formed the residue-wise confidence profile. Candidates were excluded when two or more consecutive residues within a CDR showed an absolute pLDDT difference of at least 15 points. Additional loop-specific profile-agreement criteria were applied during final candidate sorting to CDRs prioritized for structural sensitivity. Thus, filtering was based on agreement with the template pLDDT profile rather than on an absolute pLDDT threshold (**Supplementary Methods S4**).

Expression metrics and overlay filtering. Hydrophobicity was quantified using the Kyte–Doolittle scale⁵², and solvent-accessible surface area (SASA) was calculated using the Shrake–Rupley algorithm⁵³. Buried-interface and exposed non-interface hydrophobicity were derived from SASA-weighted positive hydrophathy, exposed hydrophobic residues were summarized as spatial patch-risk metrics, and VH–VL interface quality was assessed as contact count and contact density. Each candidate was then compared with quantiles of the reference antibody and developability database using predefined hydrophobicity and interface-packing criteria, with bounded rescue rules applied when a single component failed but compensatory structural features were present. Candidates passing either the base filter or an applicable rescue rule were retained through the expression-overlay step. Metric definitions, reference-database normalization, thresholds and rescue criteria are provided in **Supplementary Methods S5–S7**.

Final scoring and candidate selection. Final ranking was restricted to candidates that had passed refinement, confidence and expression-overlay filtering and that satisfied the predefined structural acceptance criteria. Candidates were required to have a global backbone RMSD of ≤ 1.5 Å, a maximum per-CDR anchor RMSD of ≤ 1.5 Å, a maximum per-CDR backbone RMSD of ≤ 2.0 Å and a maximum per-CDR side-chain RMSD of ≤ 2.5 Å. pLDDT-profile agreement further required

the absence of two or more consecutive CDR residues showing an absolute candidate–template pLDDT difference of ≥ 15 points. The mean absolute pLDDT difference was additionally required to be < 4.0 points for HCDR2 and LCDR3 and < 5.5 points for HCDR3. Final candidates were drawn from the highest-scoring set, with the expression-overlay metrics considered again at the point of selection.

Protein language-model embedding. The sequence-manifold comparison in **Extended Data Fig. 2f** used IgBERT⁵⁴. Residue-level embeddings were extracted and split into heavy- and light-chain blocks by token position, and each chain’s $L \times 1,024$ residue matrix was mean-pooled to a $1 \times 1,024$ chain-level vector before principal-component projection.

Comparator designs. CUMAb comparators²⁸ were generated for the parental antibodies of the corresponding CLade designs through the CUMAb web server. All CUMAb designs were subjected to ColabFold structure prediction for the in silico structural comparison. For the three antigen systems in which CUMAb comparators were produced experimentally (CD40, CD32a and MUC1), the comparator molecules were evaluated using the same experimental panel as the corresponding CLade designs.

In silico CDR comparison. The CDR all-atom RMSD values for CLade and CUMAb designs are reported in Supplementary Tables 4–6. For the method-level comparison in Table 1, values from multiple CLade candidates were averaged within each antigen system. The resulting CLade value and one CUMAb value per system were then averaged across 15 systems. CD5 was excluded from this comparison because its high CUMAb RMSD values would disadvantage CUMAb in the mean comparison.

Framework-anchor substitution burden. For each candidate, the 36 framework-anchor positions defined by the 3+3 anchor across all six CDRs were compared with the parental template

and the number of positions differing in residue identity was recorded as the framework-anchor substitution burden. CDR structural deviation was the sum of the six per-CDR backbone RMSDs for the same candidate. Both quantities were standardized within antigen system to remove between-system baseline differences, the Pearson correlation between them was computed separately within each system, and the direction of the system-level correlations was then summarized (**Fig. 6d; Extended Data Fig. 8a, b**).

Antibody production and purification. Heavy- and light-chain genes of the parental antibodies and of the engineered candidates were cloned separately into mammalian expression vectors. For a subset of molecules, gene synthesis, cloning, expression and purification were performed by Twist Bioscience under their standard recombinant-antibody workflow. For these externally produced antibodies, protein concentration and yield were determined by absorbance at 280 nm using a Lunatic spectrophotometer. Purity was assessed by CE-SDS using the Protein Express chip Antibody Analysis assay on a LabChip instrument. Under reducing conditions, heavy-chain and light-chain peak areas were quantified as percentages of the total detected protein. For in-house production, heavy- and light-chain genes were cloned into pcDNA3.4 and the inserts verified by Sanger sequencing. Sequence-verified plasmid pairs were co-transfected into HEK293F cells using the ExpiFectamine 293 Transfection Kit and cultured for 5 days at 37 °C, 8% CO₂ and 125 rpm. Supernatants were clarified by centrifugation, with additional filtration where required, and IgG was captured on Protein A resin, washed extensively with 1× PBS containing 0.05% Triton X-114 and eluted with 100 mM glycine, pH 2.5, directly into 1 M Tris, pH 8.0, for immediate neutralization. Purified antibody was buffer-exchanged into PBS, pH 7.4, quantified by absorbance at 280 nm on a NanoDrop spectrophotometer and stored at –80 °C until analysis.

805 **Size-exclusion chromatography.** Purified antibody (20 μ g) was injected onto a Superdex 200
806 Increase 10/300 GL column in PBS, pH 7.4, at 0.5 ml min⁻¹, with elution monitored by absorbance
807 at 280 nm. The principal peak at the retention volume expected for intact IgG monomer was
808 integrated as the monomer fraction, and monomer content was calculated as the monomer peak
809 area divided by the total integrated protein peak area $\times$ 100. SEC profiles and monomer
810 percentages were compared with those of the cognate template antibody. Each sample was
811 analysed once.

812 **Nano differential scanning fluorimetry (nanoDSF).** Thermal stability was assessed by nano
813 differential scanning fluorimetry. Antibody samples were prepared at 0.5 mg/ml in 43 mM citrate
814 and 147 mM HEPES, pH 6.0, and analysed using a temperature ramp rate of 1 $^{\circ}$ C/min. Data were
815 processed using PR.Panta Analysis software v1.10.3. Thermal stability was characterized by the
816 onset temperature of unfolding (T_{onset}) and the detected melting-transition temperatures (T_{m1} , T_{m2}).
817 Measurements were performed in triplicate.

818 **Dynamic light scattering.** Dynamic light scattering was performed at 20 $^{\circ}$ C using antibody
819 samples at 0.5 mg/ml in 43 mM citrate and 147 mM HEPES, pH 6.0. Data were processed using
820 PR.Panta Analysis software v1.10.3. The polydispersity index (PDI) and cumulant hydrodynamic
821 radius were derived from the resulting particle-size distribution profiles.

822 **AC-SINS.** AC-SINS was performed using antibody samples and the baseline control (muIgG2a)
823 at 0.05 mg/ml in 1 $\times$ PBS, pH 7.4. Self-interaction was quantified from the shift in plasmon
824 wavelength following incubation of antibody samples with capture-antibody-coated gold
825 nanoparticles. The wavelength shift ($\Delta\lambda_{\text{max}}$) was calculated as the difference between the plasmon
826 wavelength of the sample–nanoparticle conjugate and the muIgG2a negative control baseline
827 plasmon wavelength of the capture-antibody-coated nanoparticles.

828 **High-throughput hydrophobic interaction (HIC) assay.** Hydrophobic interaction was assessed
829 by Twist Bioscience using the PAIA PA-DEV-HIC assay at a nominal antibody concentration of
830 0.01 mg/ml in 1× PBS, pH 7.4. Samples were incubated with hydrophobic beads and a
831 fluorescence-labelled Fc-binding nanobody. Fluorescence was measured at 640 nm and reported
832 in relative fluorescence units (RFU). Higher RFU values indicated lower apparent hydrophobicity.

833 **Antigen-binding ELISA.** Target antigen was diluted in PBS to 100 ng/well and coated onto 96-
834 well microplates at 50 µl per well overnight at 4 °C. For assay-specific exceptions using an
835 antibody-coated format, the corresponding antibody was coated and target antigen was titrated
836 instead; the coating reagent, starting concentration and dilution scheme for each antigen system
837 are provided in Supplementary Table 7. Plates were blocked with 3% BSA for 1 h at room
838 temperature. For antigen-coated assays, antibodies were serially diluted over concentration ranges
839 selected to encompass the measurable dose-response range; starting concentrations and dilution
840 series therefore varied among antibodies. Detailed assay-specific coating conditions, titrated
841 analytes, concentration ranges and dilution schemes are provided in Supplementary Table 7.

842 Template, CUMAb and CLade molecules were compared under matched assay conditions within
843 each antigen system. After 1 h at room temperature, plates were washed with PBS-T. For antigen-
844 coated assays, bound human antibodies were detected with HRP-conjugated goat anti-human IgG
845 Fc antibody (arigo Biolaboratories, ARG23874; 1:12,000). For antibody-coated assays using GST-
846 tagged antigen, bound antigen was detected with HRP-conjugated rabbit anti-GST antibody
847 (Immunology Consultants Laboratory, RGST-45P-Z; 1:15,000). Plates were incubated with the
848 corresponding detection antibody for 1 h at room temperature, developed with TMB substrate,
849 stopped with 1 M H₂SO₄ and read at 450 nm. Dose-response curves were plotted and analysed
850 using GraphPad Prism 8. Apparent EC₅₀ values were calculated from the concentration–response

data using the Quest Graph™ EC₅₀ Calculator (AAT Bioquest) with the default calculator settings⁵⁵. EC₅₀ values were reported only when a concentration-dependent response could be resolved within the experimentally tested concentration range. Each antibody was measured in three independent experiments, and the EC₅₀, maximal signal and overall concentration-dependent profile of each candidate were compared with those of the cognate template measured under the corresponding assay conditions.

Polyreactivity. Polyreactivity was assessed by ELISA against dsDNA (1 µg/ml; prepared in house), human insulin (5 µg/ml; Sigma-Aldrich, I9278), cardiolipin (50 µg/ml; Sigma-Aldrich, C0563) and KLH (5 µg/ml; Sigma-Aldrich, H8283)⁵⁶. Antibodies were tested at 100 nM and detected using HRP-conjugated goat anti-human IgG Fc (arigo Biolaboratories, ARG23874; 1:12,000), followed by TMB development and absorbance measurement at 450 nm. Pembrolizumab, trastuzumab and adalimumab were included as reference antibodies and RT11i as a positive control. Measurements were performed in three independent experiments.

Cell lines and culture. Cell lines used for cell-surface antigen-binding assays were maintained at 37 °C in a humidified atmosphere containing 5% CO₂. Jurkat cells were cultured in RPMI-1640 medium (WELGENE, LM 011-01). T98G, MG-63, PANC-1 and FZD10-overexpressing HEK293T cells were cultured in DMEM (Cellpick, CPD002). All media were supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, A5669701) and 1× penicillin-streptomycin (Thermo Fisher Scientific, 15140122). Cells were maintained under these conditions unless otherwise specified for an individual assay.

Cell binding assay. Cell surface binding was evaluated by flow cytometry using cell lines expressing the corresponding target antigen. For each antigen system, cells were prepared at the indicated cell number and incubated with the corresponding template, reference and engineered

antibodies over the specified concentration range for 1 h at 4 °C. Cell-bound antibody was detected with PE-conjugated goat anti-human IgG Fc (Invitrogen, 12-4998-82; 1:200) for 20 min at 4 °C in the dark, followed by flow cytometric analysis on a BD FACSCanto II. The cell line, cell input and antibody concentration range used for each antigen system are provided in Supplementary Table 8.

Experimental arms and antigen system grouping. The developability-engineering arm comprises molecules in which the framework or local sequence of an antibody that already binds its antigen was redesigned to improve biophysical behaviour such as recombinant expression and purity. The developability-grafting arm comprises scaffold transfer designs in which a donor paratope was moved onto a clinically conditioned recipient framework. The 16 antigen systems are CD5, CD40, CD32a, MERTK, FZD10 A, FZD10 B, MUC1, EpCAM, GD2, B7H3, IGF1R, MSTN, IL18, IL4, LOX1 and TAG72. FZD10 A and FZD10 B were treated as separate antigen systems because they were derived from distinct parental antibodies.

Developability threshold definition. Thresholds were defined from assay-specific control measurements rather than from the CLade candidate data. For T_{onset} , T_{m1} , T_{m2} , PDI, AC-SINS and HIC they were calibrated from the Twist assay control distributions; for yield, SEC monomer content and polyreactivity, which are not represented in that panel, pembrolizumab, trastuzumab and adalimumab were expressed and evaluated as reference controls. Binding retention was defined as an apparent EC_{50} within tenfold of the original donor antibody from which the paratope was taken. Template rows in the deposited data records are normalized to 1.000 as self-references. For the anti-CD5 series, developability engineering started from an intermediate CLade clone, so binding retention is referenced to the parental antibody (apparent EC_{50} 0.210 nM; **Fig. 5f**), whereas the matched template for the developability comparison is that intermediate clone. These criteria

were applied to the assessments shown in **Fig. 6f–h and Extended Data Fig. 8d–j**; detailed threshold derivation and control distributions are provided in **Supplementary Methods S17** and **Supplementary Table 2**.

Composite developability criterion. Antigen binding retention, recombinant expression yield and SEC monomer content were defined as mandatory, whereas T_{onset} , T_{m1} , T_{m2} , PDI, AC-SINS and HIC were treated as non-mandatory supporting metrics (**Fig. 6f–h; Extended Data Fig. 8d–i**). The four polyreactivity assays were evaluated individually and combined into one additional non-mandatory metric (**Extended Data Fig. 8j**). Failure of any measured mandatory gate resulted in failure of the corresponding design. In the developability-engineering arm, a supporting developability property that did not meet the absolute threshold could be rescued when the corresponding property improved in the favourable direction relative to the matched template; this rescue was not applied to mandatory gates or to the developability-grafting arm (**Supplementary Table 1**). This approach reflects the fact that the engineering templates were selected for suboptimal biophysical properties, such that improvement relative to the starting molecule was itself a relevant engineering endpoint.

The allowable number of supporting metric failures was calibrated from the Twist assay controls using the 95% confidence-interval pass thresholds. The median number of failed supporting metrics was two per control antibody, as shown in **Supplementary Table 3**, and designs were therefore permitted up to two non-mandatory failures. Missing or invalid measurements were excluded from scoring rather than imputed or counted as failures. Designs were scored individually, and an antigen system was considered successful when at least one evaluated design satisfied the candidate-level criterion. Candidate-level measurements and scoring inputs are available as

described in the **Data Availability statement**, with further details provided in **Supplementary Methods S18**.

Numbering and region definitions. Variable domains were numbered by the Kabat scheme⁴⁴. CDR boundaries followed the Kabat definition and were assigned with AbNumber. The same numbering and boundary rule was applied to templates, candidates, comparator designs and all reference structures. Vernier-zone positions followed Foote and Winter⁸.

Statistics and the unit of analysis. CDR-anchor generalization analyses are descriptive. The comparisons in **Fig. 3b, c, e, f and Extended Data Fig. 4b–f and Extended Data Fig. 5** are pooled across loops or anchor groups, so observations contributed by the same antibody are not independent. Inferences drawn at the antibody level were instead tested on dataset-level statistics, namely one-sample or paired tests across the 35 datasets for the anchor-burden and angle relationships (**Fig. 3d; Extended Data Fig. 4a**) and across antigen systems for the experimental candidates (**Fig. 6d; Extended Data Fig. 8a, b**). Zero differences were excluded from Wilcoxon signed-rank tests but retained in the reported dataset. For **Fig. 3f**, leave-one-structure-out recalculation was used to assess robustness of the observed direction rather than for statistical inference, and no P value was assigned to this analysis. For the alignment-ranking benchmark, the antibody dataset was the statistical unit because Anchor and Global alignment were applied to the same structural inputs. Dataset-level paired differences were evaluated using two-sided Wilcoxon signed-rank tests, with P values across cutoff-specific comparisons adjusted using the Holm method. Experimental developability measurements were summarized at the candidate level, whereas composite success was summarized both at the antigen-system level across the 16 evaluated systems and at the candidate level across the 23 evaluated designs. Missing or invalid measurements were excluded from pass-fail scoring rather than imputed or counted as failures.

Candidate-level experimental measurements, scoring inputs and excluded measurements are provided in the data records described in the **Data Availability statement**.

Data availability

All data needed to evaluate the conclusions of this study have been deposited in Figshare under the reserved DOI 10.6084/m9.figshare.34086867. The deposit contains the 23 final candidate sequences with their cognate templates and comparators, the per-design experimental measurements for every assay class including replicate-level values, the V-germline pairing frequency table and the composition of the clinical reference set that together define the clinical prior (Fig. 2), and the anchor definitions, junction angles and CDR RMSD measurements underlying Figs. 3 and 4. Source data for every main and Extended Data figure panel are included in the same deposit. The deposit is available to referees during review and will be released publicly on publication.

Code availability

Code for the CLade pipeline is available at github.com/Theleelab-antibody/CLade under an academic-use licence, and will be made public and archived with a DOI at the time of publication. The repository provides the anchor extraction, junction-angle computation, candidate-ranking and final-scoring code together with the clinical reference set required to run it. The scripts used to assemble that reference set from Thera-SAbDab are not included, and the assembly procedure is described in Methods. Access for referees during review is provided on request to the corresponding author.

Methods references

47. Raybould, M. I. J. *et al.* Thera-SAbDab: the Therapeutic Structural Antibody Database. *Nucleic Acids Res.* **48**, D383–D388 (2020).
48. Giudicelli, V., Chaume, D. & Lefranc, M.-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* **33**, D256–261 (2005).
49. Dunbar, J. & Deane, C. M. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**, 298–300 (2016).
50. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
51. Kallenborn, F. *et al.* GPU-accelerated homology search with MMseqs2. *Nat. Methods* **22**, 2024–2027 (2025).
52. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
53. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371 (1973).
54. Kenlay, H. *et al.* Large scale paired antibody language models. *PLOS Comput. Biol.* **20**, e1012646 (2024).
55. AAT Bioquest, Inc. Quest GraphTM EC50 Calculator. <https://www.aatbio.com/tools/ec50-calculator> (2026).
56. Lim, J.-Y. *et al.* Normalized internalization predicts potency of carcinoembryonic antigen-related cell adhesion molecule 6-targeting antibody-drug conjugates. *J. Controlled Release* **396**, 115106 (2026).

985

986 **Acknowledgements**

987 This work was supported in part by the Bio&Medical Technology Development Program of the
988 National Research Foundation (NRF) through the Innovative New Drug Development Program
989 (No. RS-2026-25527526) and the Mid-Career Researcher Program (No. RS-2023-00278980), both
990 funded by MSIT.

991 **Author contributions**

992 C.-H.L. conceived the study and directed the overall research strategy. W.J. initiated the project
993 and, together with M.P., designed and implemented the computational pipeline, performed the
994 computational analyses, generated the figures and wrote the original draft of the manuscript. M.P.
995 and J.H.C. performed the SEC experiments. J.H.C. coordinated the experimental work and
996 performed the binding and polyreactivity assays. H.L., M.M.J.K., J.H., Y.K., J.Y., J.S., Y.L. and
997 Y.J. performed the antibody characterization experiments, each on a distinct subset of the antibody
998 panel. C.-H.L. guided the design and interpretation of all experiments and computational analyses,
999 supervised the project, provided resources and acquired funding. All authors discussed the results
1000 and commented on the manuscript.

1001 **Competing interests**

1002 C.-H.L., W.J. and M.P. are listed as inventors on Korean patent application no. 10-2025-0013450,
1003 which covers the automated antibody stability engineering and humanization software described
1004 in this work. The remaining authors declare no competing interests.

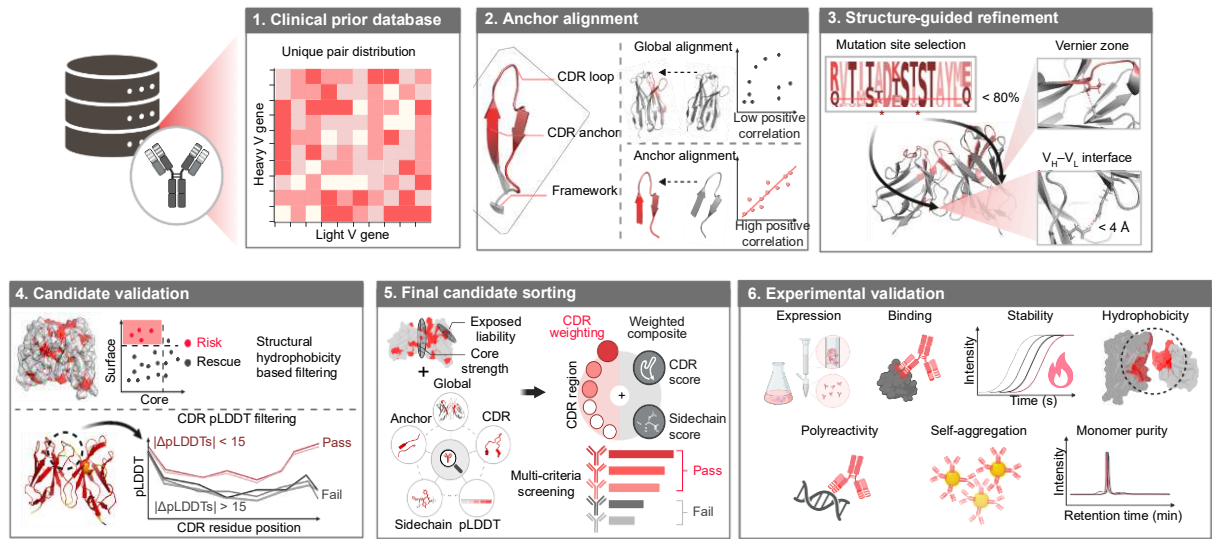
1005 **Additional information**

1006 Supplementary Information is available for this paper. Correspondence and requests for materials

1007 should be addressed to Chang-Han Lee.

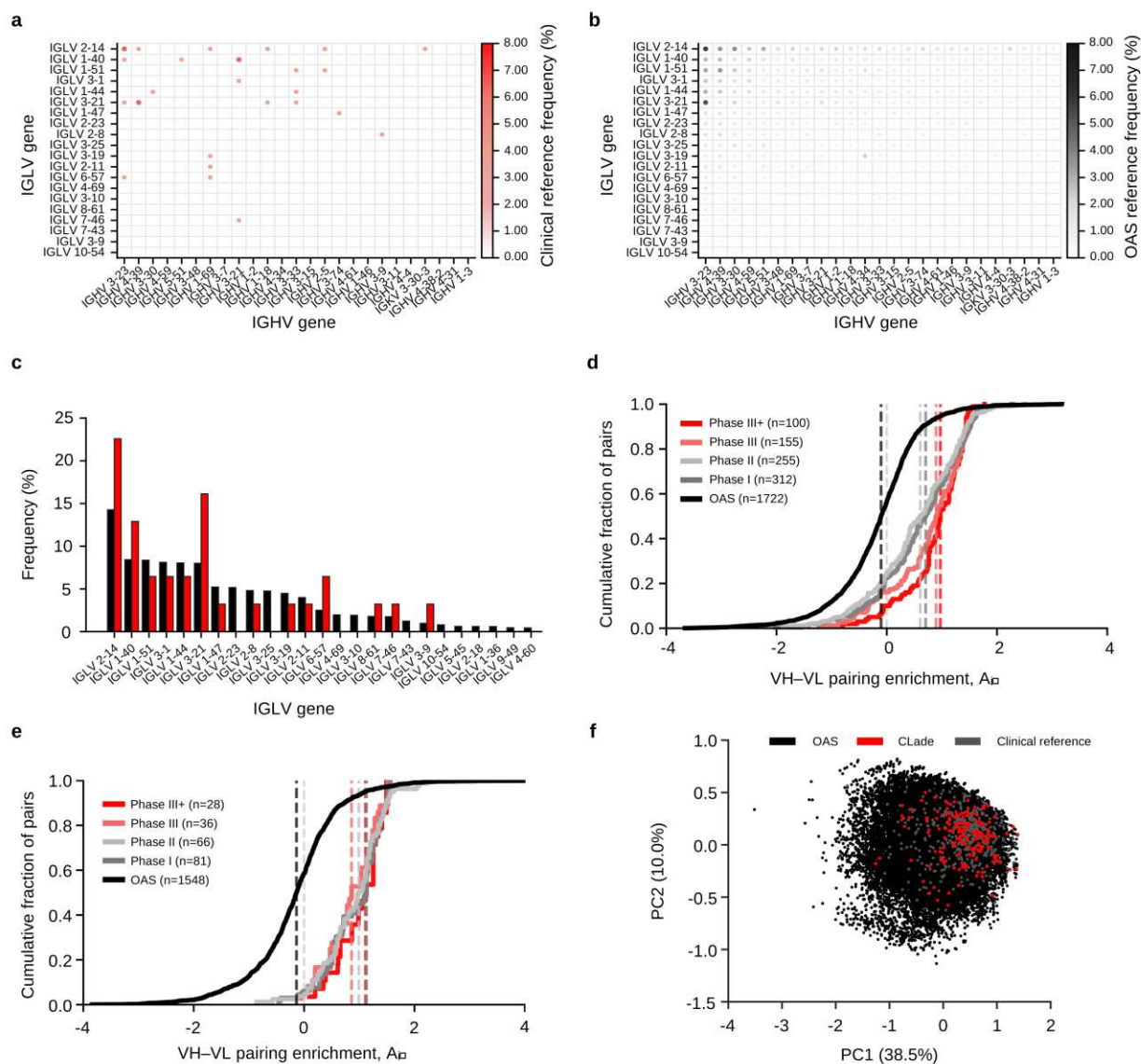
1008

Extended Data figure legends



Extended Data Fig. 1. Overview of the CLade workflow: five computational operations followed by experimental validation.

Construction of a clinical-prior framework database, local CDR-anchor alignment, structure-guided refinement at conserved Vernier zone and V_H-V_L interface positions, structural-confidence and hydrophobicity-based candidate validation, multi-criterion candidate sorting, and experimental validation.



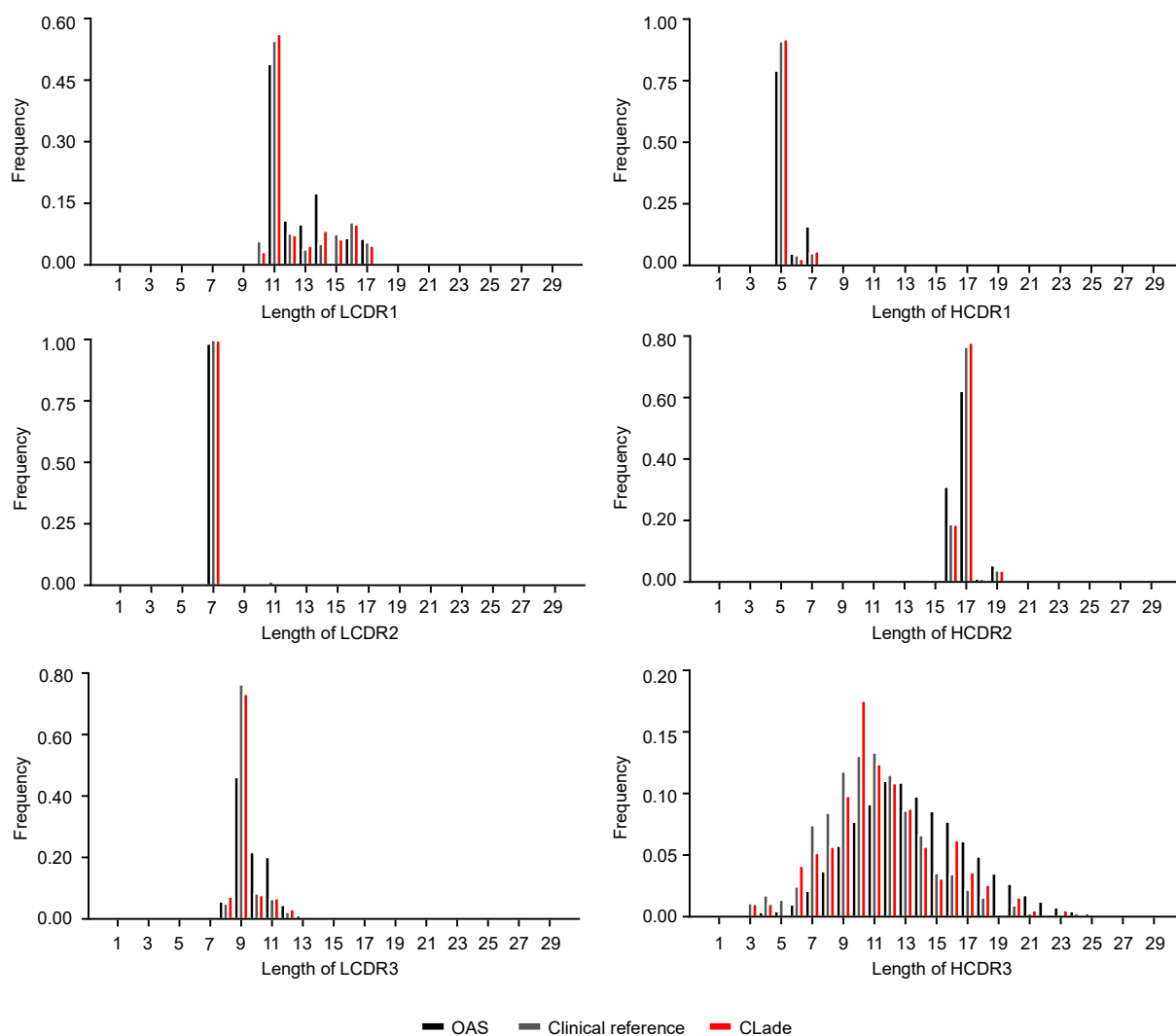
Extended Data Fig. 2. Lambda-chain, sequence-manifold, and pair-enrichment views of the clinical prior.

(a) VH-Vλ germline-pairing landscape in the clinical reference set. Bubble size and colour intensity indicate pair frequency. **(b)** VH-Vλ germline-pairing landscape in OAS, displayed on the same axes as in **(a)**. **(c)** IGLV gene-usage frequencies in OAS (black) and the clinical reference set (red). **(d)** ECDF of VH-Vκ pairing enrichment across cumulative clinical stages, quantified using pseudocount-stabilized A_{ij} . Positive values indicate enrichment above that expected from

1026 marginal VH and V κ frequencies. **(e)** ECDF of VH–V λ pairing enrichment displayed as in **(d)**. **(f)**
1027 Protein language-model embedding projection of germline sequences from OAS, the CLade
1028 database, and clinical reference antibodies. Points are coloured by dataset. PC1 and PC2 denote
1029 the two principal components of the sequence embeddings.

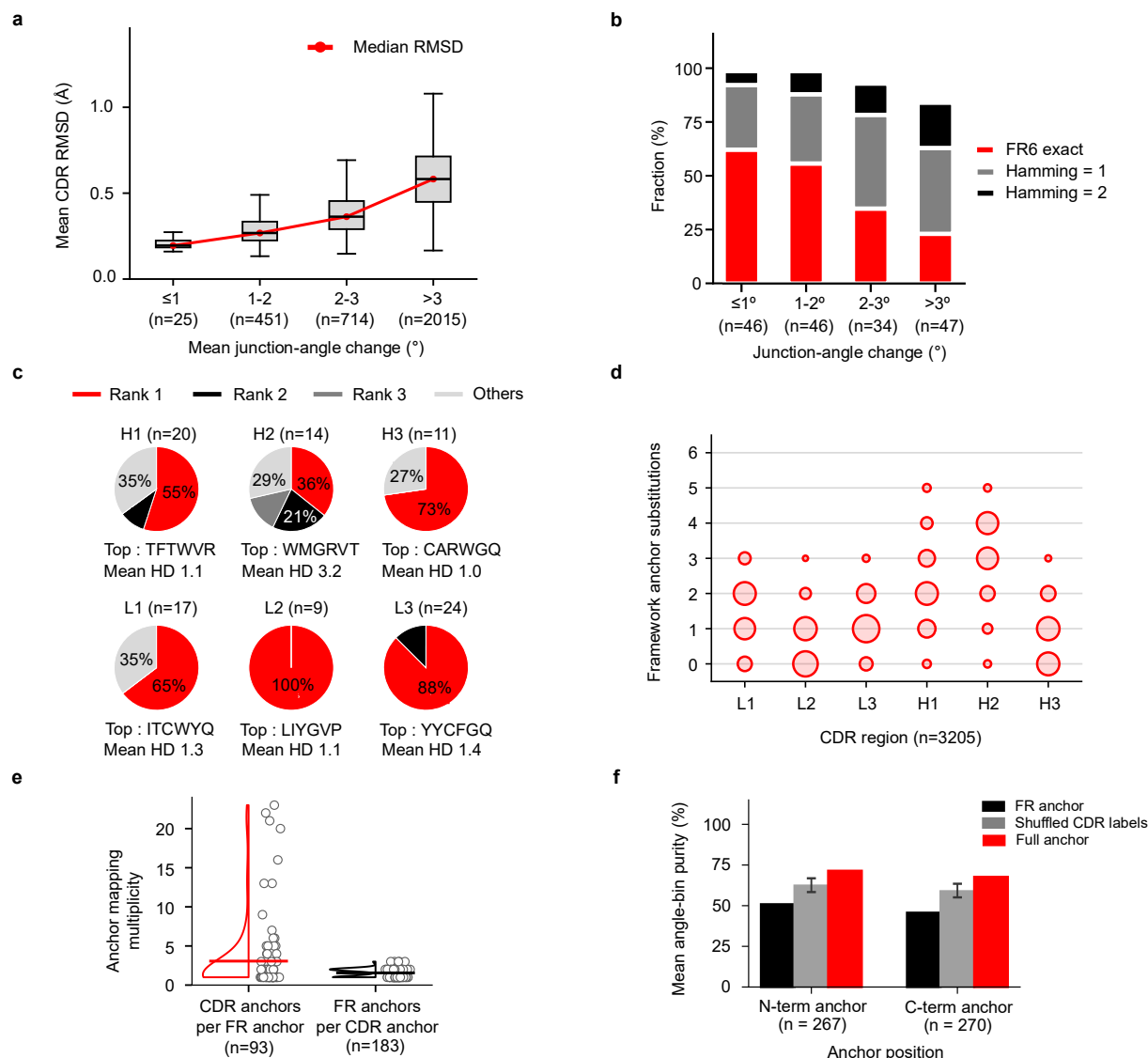
1030

1031



Extended Data Fig. 3. CDR loop-length distributions across OAS, CLade, and the clinical reference set.

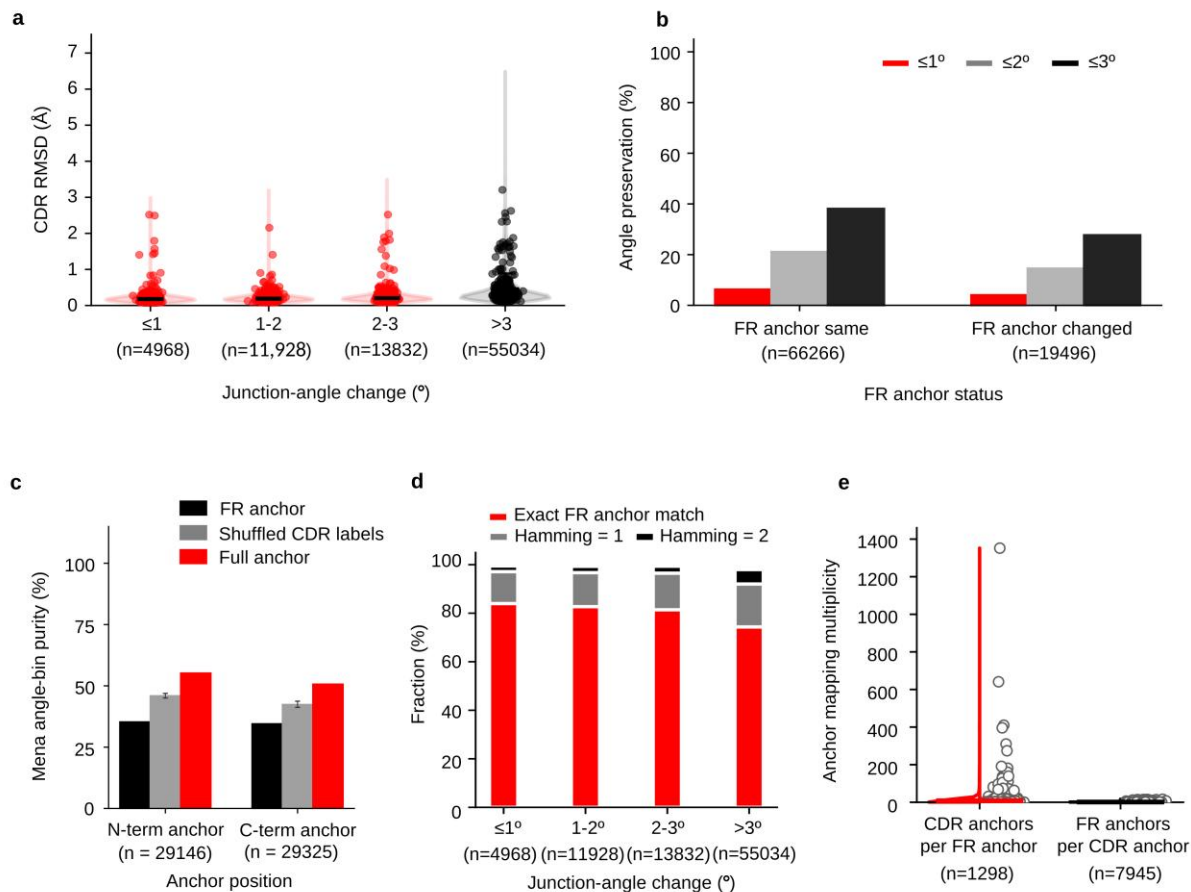
Length distributions of HCDR1, HCDR2, HCDR3, LCDR1, LCDR2, and LCDR3 in OAS (black), the CLade database (red), and the clinical reference set (grey).



Extended Data Fig. 4. Anchor specificity, mapping multiplicity, and large-scale scaffold-ranking analyses.

(a) Integrated CDR RMSD, defined as the arithmetic mean of the six CDR-specific RMSDs, across candidate mean junction-angle-change bins. **(b)** Distribution of exact framework-anchor (FR6) matches and FR6 sequences at Hamming distances of 1 or 2 across maximum junction-angle-change bins. The unfilled remainder represents Hamming distance >2. **(c)** Rank distribution of recurrent framework-anchor motifs among cases with altered framework anchors for HCDR1–

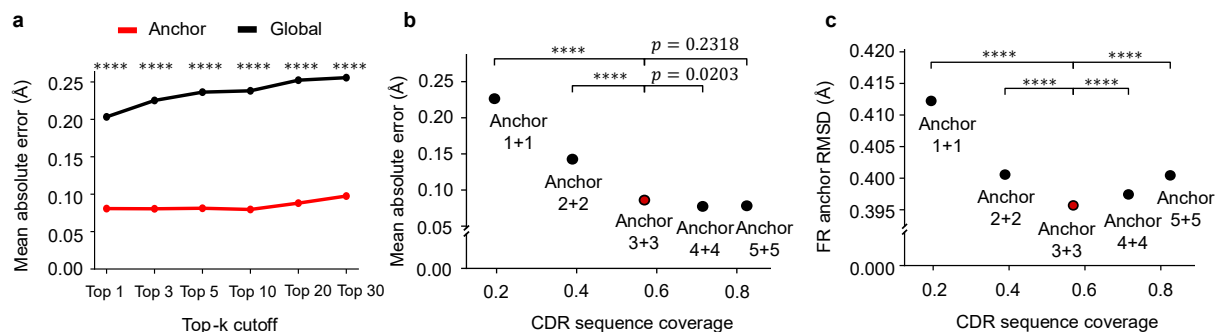
1046 HCDR3 and LCDR1–LCDR3. Pie sectors indicate the first-, second-, and third-ranked motifs and
1047 all others. The dominant motif and mean Hamming distance from the parental anchor are shown
1048 below each plot. **(d)** Loop-specific distribution of framework-anchor substitution counts among
1049 3,205 candidates. For each CDR, substitutions among the six framework-anchor residues are
1050 counted from 0 to 6. Bubble area is proportional to the fraction of candidates in each category. **(e)**
1051 Bidirectional mapping multiplicity between framework and CDR anchors. Each point is an anchor-
1052 sequence group. Distributions show the number of distinct CDR anchors per framework anchor (n
1053 = 93 groups) and distinct framework anchors per CDR anchor (n = 183 groups). **(f)** Fraction of
1054 observations assigned to the dominant 3-degree junction-angle bin when conditioned on the
1055 framework anchor alone, the framework anchor with shuffled CDR-anchor labels, or the full
1056 anchor. Error bars on the shuffled-label reference show the central 95% of the permutation
1057 distribution.
1058



Extended Data Fig. 5. Large-scale structural replication of anchor-geometry relationships.

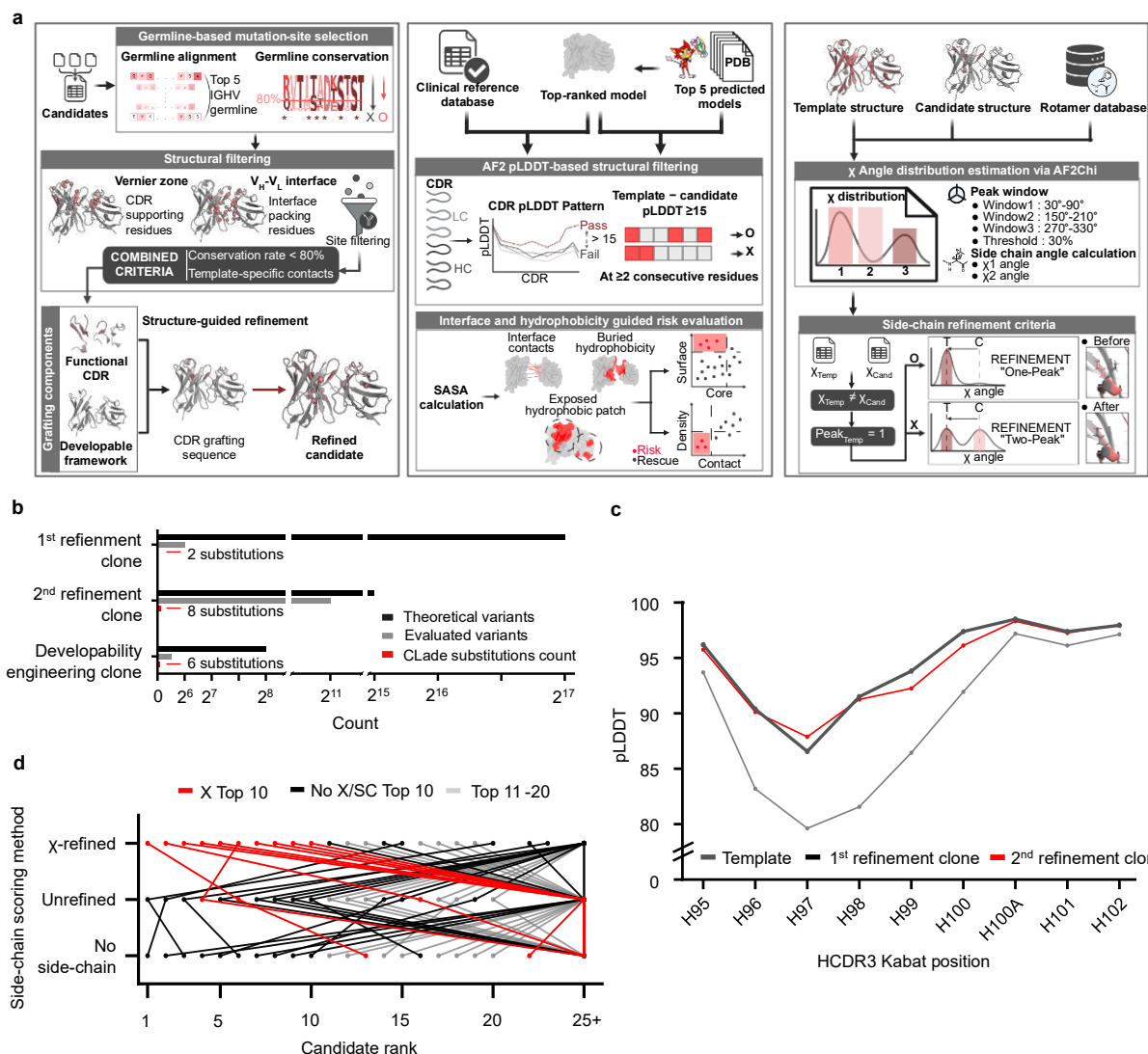
(a) CDR backbone RMSD across bins defined by the larger of the N-terminal and C-terminal junction-angle changes in the SAbDab structure-pair dataset. Bin sizes are $n = 4,968$, $11,928$, $13,832$, and $55,034$ for $\leq 1^\circ$, $>1^\circ$ to $\leq 2^\circ$, $>2^\circ$ to $\leq 3^\circ$, and $>3^\circ$, respectively. **(b)** Junction-angle preservation for CDR-identical pairs with the framework anchor unchanged ($n = 66,266$) or changed ($n = 19,496$). Bars show the fractions within $\leq 1^\circ$, $\leq 2^\circ$, and $\leq 3^\circ$. **(c)** Dominant 3-degree junction-angle-bin occupancy after conditioning on the framework anchor alone, shuffled CDR-anchor labels, or the full anchor at the N-terminal ($n = 29,146$) and C-terminal ($n = 29,325$) junctions. Error bars show the central 95% of the permutation distribution. **(d)** Distribution of exact FR6 matches and FR6 sequences at Hamming distances of 1 or 2 across angle-change bins in the

1070 expanded dataset. The unfilled remainder represents Hamming distance >2. **(e)** Bidirectional
1071 mapping multiplicity in the expanded dataset, shown as distinct CDR anchors per framework
1072 anchor (n = 1,298 groups) and distinct framework anchors per CDR anchor (n = 7,945 groups).
1073
1074



Extended Data Fig. 6. Alignment error and anchor-window selection, and a representative ranking transition.

(a) Dataset-mean absolute deviation from the oracle CDR reference for anchor and global alignment across top-k candidate sets. The analysis unit is the dataset. P values are from two-sided paired Wilcoxon signed-rank tests with Holm correction across the six cutoffs. **(b)** Mean absolute error from the oracle CDR reference versus direct CDR-sequence coverage for 1+1, 2+2, 3+3, 4+4, and 5+5 framework-anchor windows. Red denotes the selected 3+3 anchor. P values compare each window with 3+3 by two-sided paired Wilcoxon signed-rank tests across 35 datasets, with Holm correction within the metric. For comparisons of 3+3 with 1+1, 2+2, 4+4 and 5+5, respectively, $W = 0, 12, 158$ and 242 ; Holm-adjusted $P = 9.91 \times 10^{-7}, 2.08 \times 10^{-6}, 0.0203$ and 0.232 ($n = 35$ datasets). **(c)** Local anchor-to-framework RMSD versus CDR-sequence coverage for the same five anchor windows. The 3+3 window yields the lowest local framework RMSD. Statistical comparisons are as in **(b)**. For comparisons of 3+3 with 1+1, 2+2, 4+4 and 5+5, respectively, $W = 30, 32, 80$ and 10 ; Holm-adjusted $P = 9.12 \times 10^{-6}, 9.12 \times 10^{-6}, 1.19 \times 10^{-4}$ and 2.35×10^{-6} ($n = 35$ datasets).

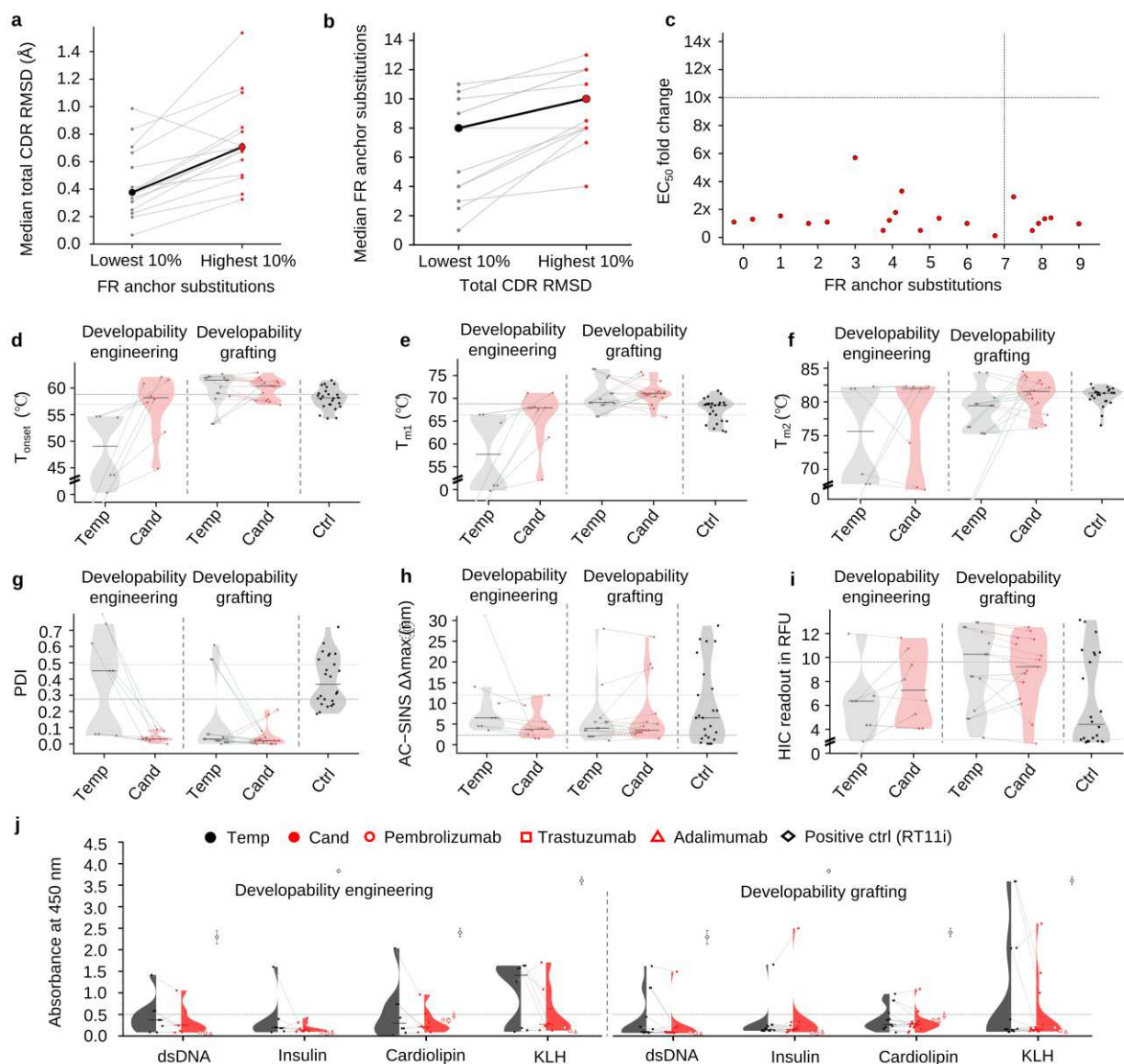


Extended Data Fig. 7. CLade refinement, confidence filtering, substitution space restriction, and chi-angle-aware ranking.

(a) Integrated refinement and candidate-filtering workflow. Germline conservation identifies candidate substitution positions, which are restricted to Vernier-zone and VH-VL interface residues. AF2 χ -based chi-angle distributions guide one-peak or two-peak side-chain refinement. Candidate structures are then filtered using inter-model CDR pLDDT patterns, buried hydrophobicity, and exposed hydrophobic-patch risk. **(b)** Substitution search-space restriction for

1100 the first-refinement, second-refinement, and developability-engineering stages. Black bars show
1101 the theoretical combination space, grey bars the combinations evaluated, and red marks the
1102 substitutions retained by CLade. The final designs contain 2, 8 and 6 targeted substitutions selected
1103 from 17, 15 and 8 candidate positions, respectively. **(c)** Residue-wise pLDDT across HCDR3
1104 (Kabat H95–H102) for the template, first-refinement clone, and second-refinement clone. The
1105 second refinement restores the local confidence profile around H97–H99 toward the template. **(d)**
1106 Ranking transitions among 1,372 candidates scored without a side-chain term, with side-chain
1107 RMSD from models lacking chi-angle refinement, or with template-chi-refined side chains.

1108

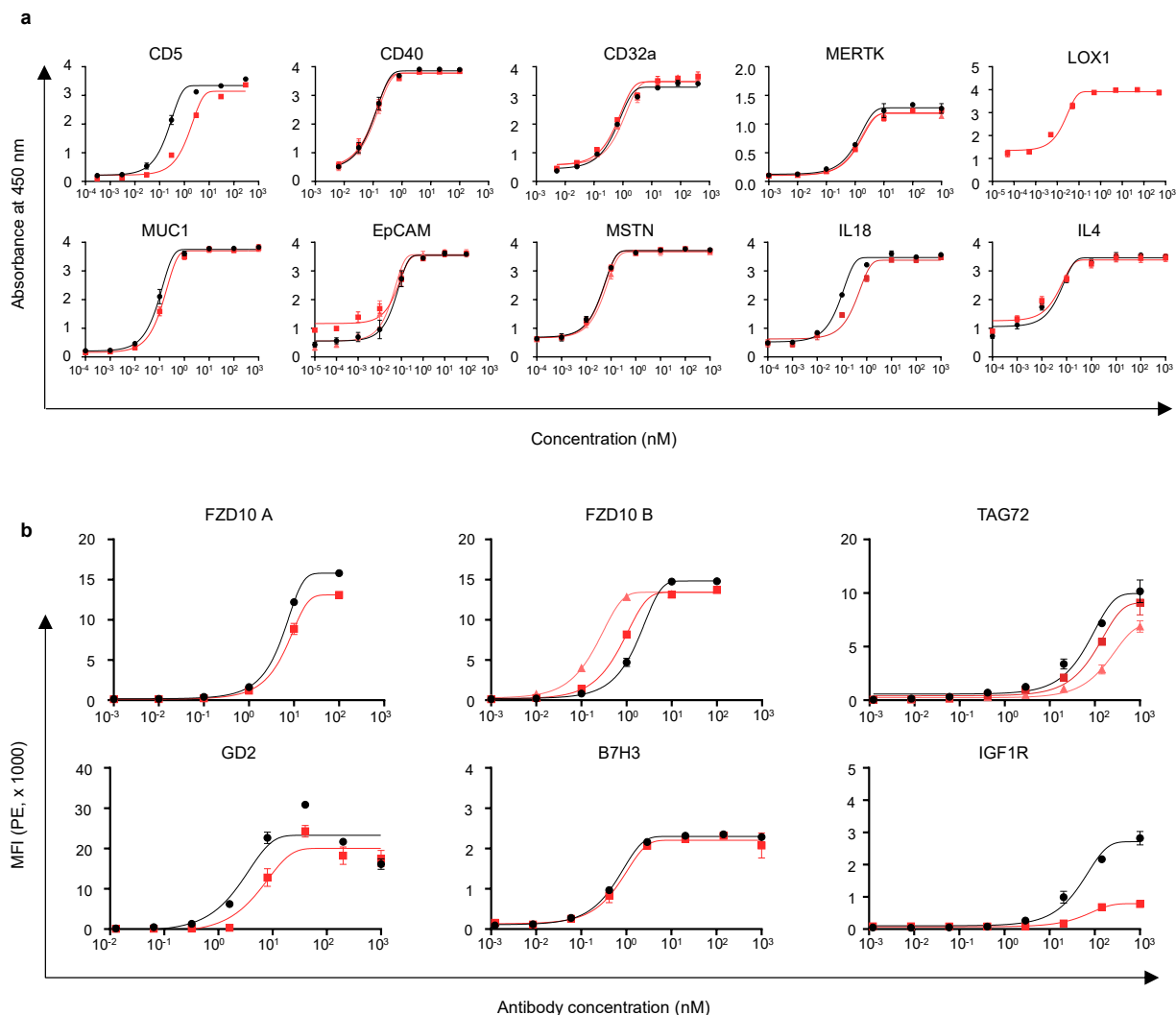


Extended Data Fig. 8. Additional developability analyses across evaluated CLade candidates.

(a) CDR backbone RMSD in the lowest and highest 10% of framework-anchor burden among germline-only grafts. Grey lines pair systems; black points and line show medians. Medians are 0.376 and 0.706 Å; 14 of 15 systems increase (median difference 0.296 Å; two-sided paired Wilcoxon signed-rank test, $W = 6$, $n = 15$, $P = 8.54 \times 10^{-4}$). **(b)** Framework-anchor substitution counts in the best and worst 10% by CDR RMSD. Grey lines pair systems; black points and line show medians. Medians are 8.0 and 10.0 substitutions; 14 of 15 systems increase, with a median

1117 paired difference of 2.5 substitutions (two-sided paired Wilcoxon signed-rank test, $W = 0$, $n = 15$,
1118 $P = 9.45 \times 10^{-4}$). **(c)** Apparent EC_{50} fold change versus framework-anchor substitutions in 20
1119 evaluable candidates. Dashed lines mark seven anchor substitutions and the 10-fold binding-
1120 retention criterion. **(d)** Paired thermal-onset temperature by arm for template (Temp) and candidate
1121 (Cand); unrelated controls are black. Good and intermediate cutoffs are ≥ 58.825 and ≥ 56.930 °C.
1122 **(e)** Paired first thermal transition (T_{m1}) for the same groups. Good and intermediate cutoffs are
1123 ≥ 68.810 and ≥ 66.390 °C. **(f)** Paired second thermal transition temperature (T_{m2}) for the 21
1124 candidates with a resolvable measurement. Good and intermediate cutoffs are ≥ 81.490 and
1125 ≥ 80.560 °C. **(g)** Paired polydispersity index (PDI); good and intermediate cutoffs are ≤ 0.275 and
1126 ≤ 0.490 . **(h)** Paired AC-SINS wavelength shift; good and intermediate cutoffs are ≤ 2.25 and ≤ 12
1127 nm. **(i)** Paired HIC readout; good and intermediate cutoffs are ≥ 9.6175 and ≥ 3.1355 RFU (higher
1128 is favourable). **(j)** Polyreactivity at 100 nM to dsDNA, insulin, cardiolipin and keyhole limpet
1129 hemocyanin (KLH) in both arms. Controls are pembrolizumab, trastuzumab, adalimumab and
1130 RT11i; the dashed cutoff is 0.5.

1131



Extended Data Fig. 9. Binding-response curves for the evaluated CLade candidates.

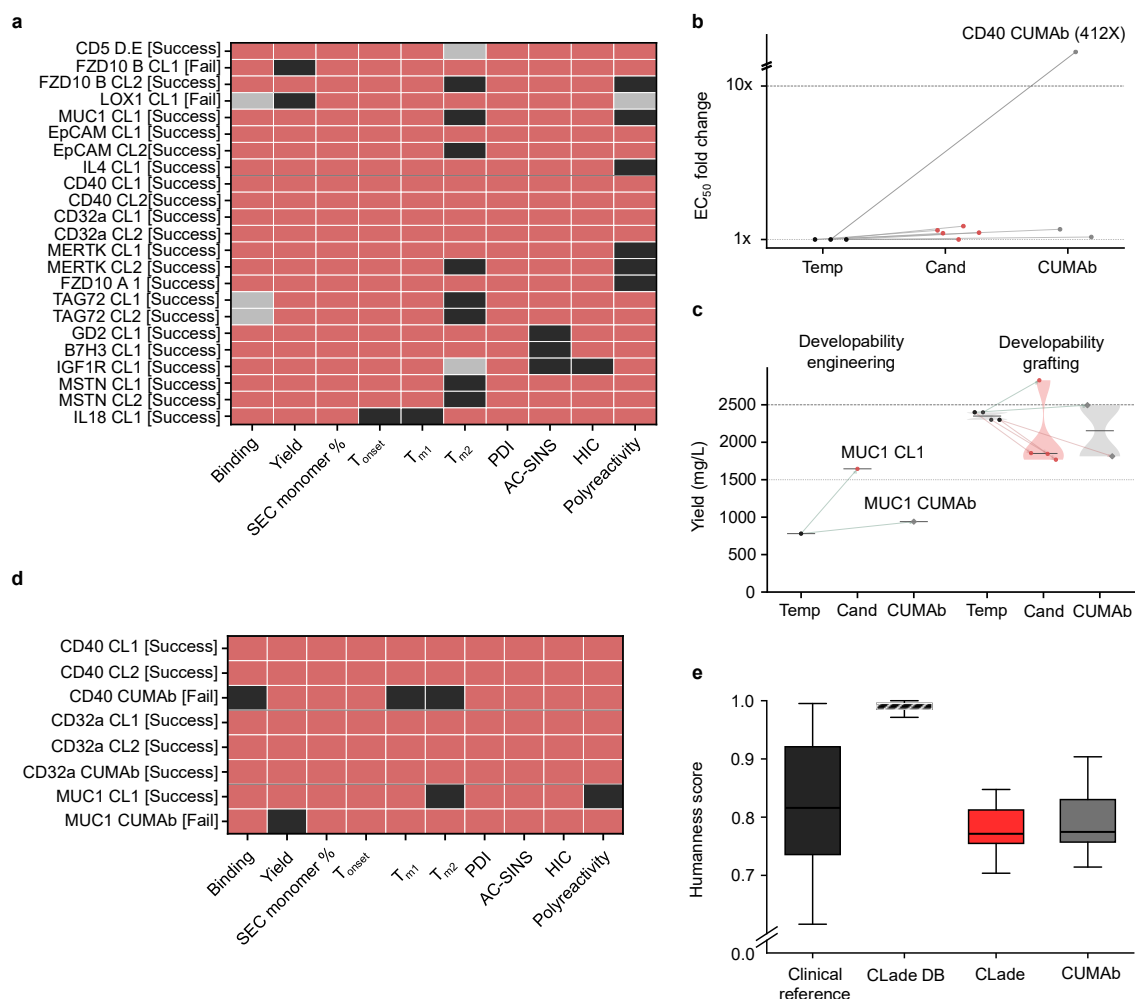
(a) Antigen-binding dose-response curves for the template (black) and CLade candidate or candidates (red) for CD5, CD40, CD32a, MERTK, LOX1, MUC1, EpCAM, MSTN, IL18, and IL4. Binding is shown as absorbance at 450 nm against the concentration indicated on each plot.

(b) PE fluorescence-based binding dose-response curves for the template (black) and CLade candidate or candidates (red) for FZD10 A, FZD10 B, TAG72, GD2, B7H3, and IGF1R. Binding

1139 is shown as mean fluorescence intensity (MFI) against antibody concentration. Points show the
1140 measured responses and lines show fitted dose-response curves.

1141

1142



Extended Data Fig. 10. Structural, sequence, and comparator analyses of the evaluated CLade candidates.

(a) Candidate-level outcome matrix for the 23-candidate cohort across binding, yield, SEC monomer content, T_{onset} , T_{m1} , T_{m2} , PDI, AC-SINS, HIC, and combined polyreactivity. Red denotes metric success, black failure, and grey not measured. Bracketed row labels show the final candidate-level outcome under the composite rule used in **Fig. 6i**. **(b)** Head-to-head binding comparison for CD40, CD32a, and MUC1 designs. Apparent EC₅₀ fold changes are shown for template, CLade candidates, and target-matched CUMAb comparators. Lower values indicate

better retention and the dashed line marks the 10-fold criterion. The CD40 CUMAb design has a 412-fold change. **(c)** Head-to-head expression-yield comparison for the same target-matched CLade and CUMAb designs. Horizontal lines indicate the good and intermediate yield criteria. The MUC1 CUMAb design falls below the mandatory intermediate gate. **(d)** Metric-level outcome matrix for five CLade candidates and three target-matched CUMAb designs across the same 10 developability metrics. Red denotes success, black failure, and grey not measured. Bracketed labels show the final composite outcome under the arm-specific rule used in **Fig. 6i**. Under the uniform absolute rule MUC1 CL1 is also scored as a failure. Cells are shown after the arm-specific rescue. T_{onset} and T_{m1} for MUC1 CL1 are below their absolute thresholds but improve over their templates. **(e)** Humanness-score distributions for Clinical reference antibodies, the CLade database, experimentally evaluated CLade candidates, and CUMAb comparators.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CLadeNatureformatsupplementaryFIN.docx](#)