EPIFORGE: A RISK-AVERSE VERIFIER-GUIDED AGENT FOR EPITOPE-CONDITIONED ANTIBODY DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Epitope-conditioned antibody design is dominated by open-loop pipelines that generate candidates, score them once with a structure predictor, and discard whatever fails, so a rejected candidate never informs the next proposal, developability is inspected only after the fact, and a design that pleases one folding oracle need not survive another. We present EPIFORGE, a closed-loop editing agent that reframes the task as verified repair of an initialized variable region. An epitope-conditioned initializer writes CDR sequences onto a fixed framework by masked discrete diffusion over a chemistry-aware coarse-grained graph. A reasoning policy then proposes edits as a short integer genotype; a parameter-free mapper projects that genotype onto a legal mutation set by sequential modular arithmetic, so no proposal can be malformed, out of bounds, duplicated or vacuous; an uncertainty-aware world model predicts the resulting interface metrics with its own variance; and a confidence-bounded selector spends the scarce verification budget where it is most informative. Acceptance is not a learned gate: every retained edit is confirmed by a real structure prediction and the metrics computed from it, and anything that does not improve the verified state is rolled back. Selection is optimistic and credit assignment conservative, since the selector adds the surrogate's predicted standard deviation while the group-relative objective subtracts it. On a SAbDab split that removes redundancy on the antibody, antigen-sequence and antigen-structure sides at once, EPIFORGE reaches 45.3% CDR-H3 recovery and 55.1% all-CDR recovery, improves CDR-H3 backbone accuracy under reporting predictors never used inside the loop, and attains the best humanness of every method compared without optimizing for it. A held-out case study recovers pose, loop conformation and sequence together at DockQ 0.649 against 0.387 for the strongest baseline, while our ablation and two further cases show these three quantities coming apart; we report where the loop trades sequence recovery for interface quality, and where it does not lead. Our code is at `https://anonymous.4open.science/r/EpiForge-82C6/`.

1 INTRODUCTION

An antibody is useful only if it binds the intended epitope, folds, and survives development. Learned design has progressed fastest on the first of these, through antigen-conditioned co-design models and backbone diffusion pipelines that place a loop against a specified surface (Luo et al., 2022; Kong et al., 2023; Watson et al., 2023; Bennett et al., 2026). Almost all of these systems share one shape: a generator produces candidates, a structure predictor scores them once, and whatever fails is discarded. Three consequences follow. Information in a rejected candidate never reaches the next proposal, so a pipeline that discards ninety-nine designs out of a hundred also discards ninety-nine labelled observations of what that epitope rejects. Developability and humanness are inspected after generation, so a design can be optimized into a region a real program would discard and the filter that catches it arrives too late to steer anything. And because the filter is a single predictor, a design tuned against it need not survive a second one, the failure that makes *in silico* antibody results hard to trust.

EPIFORGE treats epitope-conditioned design as *verified repair* rather than one-pass generation. A candidate variable region is initialized on a fixed framework, and a closed loop then proposes an edit, predicts its effect cheaply, verifies the most promising candidates with a real structure prediction, and either keeps or rolls back (fig. 1). SKETCH is a reasoning policy emitting a short integer genotype; GAUGE is a parameter-free mapper decoding it into a legal mutation set by sequential modular

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

arithmetic, so every proposal is in bounds, free of repeated positions and free of identity substitutions by construction; ECHO predicts the resulting interface metrics with its own variance and without a structure prediction; and SCOUT ranks candidates by an upper confidence bound. Two properties address the failures above. Acceptance is never a learned judgement: only a real prediction of the edited complex and the deterministic metrics computed from it can advance the state, so a policy that learns to please a surrogate cannot thereby advance its own. And selection and credit use the surrogate's uncertainty with opposite signs, so an edit the world model cannot place is worth verifying but not worth rewarding until it has been.

This decomposition fits antibody design because of an asymmetry in cost and in trust. A complex prediction followed by deterministic contact and DockQ routines costs minutes per candidate but returns the same answer every time, whereas a learned estimate of the same quantity is nearly free and is exactly the signal a policy trained against it will learn to exploit. Giving only the expensive one authority over the state turns reward hacking into a budgeting problem: the surrogate decides where calls are spent, the verifier decides what survives. Bounded edits also make a step attributable, since a parent and a child differing in one to three point mutations are close enough for a cheap sequence model to compare.

We are explicit about limits. The loop optimizes *one* structural verifier, so the third failure is measured rather than assumed away: the reporting panel is disjoint from that verifier and any gain that fails to transfer is visible in our tables. Developability and humanness are measured, never optimized, making table 2 a check the loop may fail, and our split removes redundancy on the CDR-H3, antigen-sequence and antigen-structure sides at once. What the design does not deliver: it is level with the strongest baseline on epitope recall, its advantage on affinity maturation is unestablished, and in a controlled ablation the loop *lowers* sequence recovery while improving interface quality, a dissociation visible in aggregate (table 3) and on individual complexes (section 4.5) that we report rather than bury.

Our contributions are as follows: (1) A closed-loop formulation of epitope-conditioned design as verified repair, in which only a structural verifier and the deterministic metrics computed from it may advance the state, with two elementary statements fixing what the learned components can and cannot do. (2) A parameter-free action decoder projecting an arbitrary integer genotype onto a legal mutation set by sequential modular arithmetic, so malformed, out-of-range, position-colliding and vacuous proposals are impossible by construction while the map stays surjective onto all bounded edits. (3) A learning rule that uses the surrogate's uncertainty with opposite signs in selection and in credit, charging exploration to the verification budget rather than to the policy, and an evaluation whose reporting panel is disjoint from the in-loop verifier and whose split controls antibody, antigen-sequence and antigen-structure redundancy at once.

2 RELATED WORK

Antibody design and its open loop. Learned antibody design began with iterative-refinement graph models (Jin et al., 2021; 2022) and moved to antigen-conditioned diffusion over residue types, positions and orientations (Luo et al., 2022), then to equivariant and full-atom settings (Kong et al., 2022; 2023) and to designs guided by evolutionary, physical or geometric constraints (Verma et al., 2023; Zhu et al., 2024; Lin et al., 2024). Fixed-backbone designers remain the workhorses (Dauparas et al., 2022; Hsu et al., 2022; Høie et al., 2025), antibody language models supply sequence priors and humanness signals (Ruffolo et al., 2021; Olsen et al., 2024; 2022; Shuai et al., 2023), and backbone diffusion followed by fixed-backbone design is the dominant *de novo* route for epitope-targeted antibodies (Watson et al., 2023; Bennett et al., 2026; Ingraham et al., 2023). All are open-loop with respect to their filter. EPIFORGE uses such a pipeline as an initializer and adds the missing feedback path, so the filter becomes a reward rather than a sieve; and unlike Jin et al. (2021), which refines inside one forward model, every EPIFORGE edit is externally confirmed or rolled back.

Evaluation, developability and learning machinery. Therapeutic viability is assessed through interpretable descriptors (Raybould et al., 2019) and repertoire-derived humanness (Prihoda et al., 2022; Olsen et al., 2022), applied after generation by existing systems; we keep them there deliberately, as post hoc measurements from the sequence and a fast antibody-only predicted structure (Kenlay et al., 2024; Ruffolo et al., 2023), so they remain a test the loop can fail. Complex prediction is the standard *in silico* evaluator (Evans et al., 2021; Mirdita et al., 2022; Jumper et al., 2021; Abramson et al., 2024; Passaro et al., 2025; team et al., 2024; Team et al., 2025), with interface quality measured

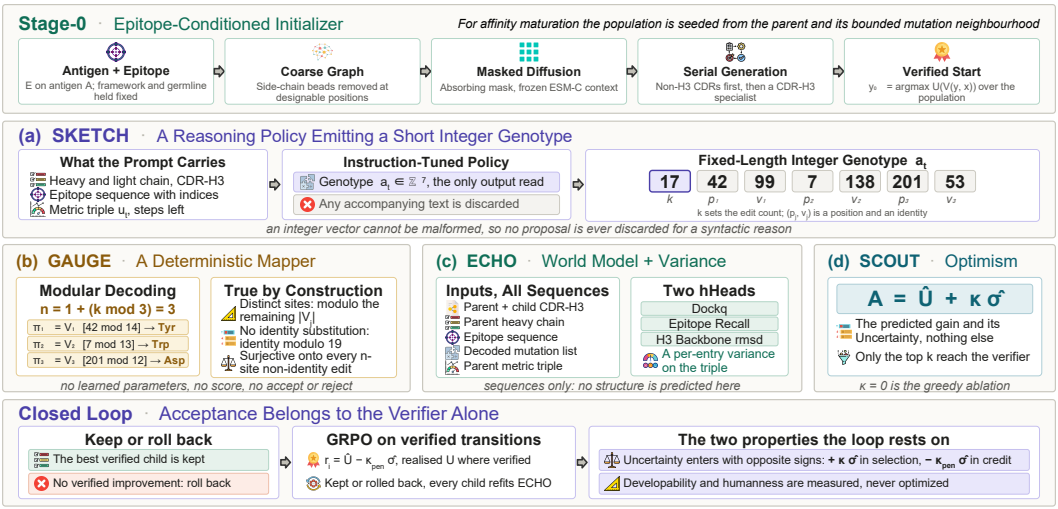

Figure 1: **EPIFORGE as verified repair.** Stage-0 initializes the variable region and its best *verified* member becomes y_0. One editing step runs left to right: (a) SKETCH emits an integer genotype, the only part of its output the pipeline reads; (b) GAUGE projects it onto a legal mutation set by modular arithmetic, with no parameters and no accept-or-reject decision; (c) ECHO predicts the child metric triple from sequences alone with a per-entry variance; (d) SCOUT forwards the top k to $\mathcal{V}$. Acceptance belongs to $\mathcal{V}$ alone, and the surrogate's uncertainty enters selection and credit with opposite signs.

by established metrics (Basu & Wallner, 2016; Lensink et al., 2007); because predictors disagree, our reporting panel is disjoint from the in-loop verifier and each predictor is reported separately rather than pooled. On the learning side, preference optimization against energy-derived labels improves antibody rationality (Zhou et al., 2024; Rafailov et al., 2023), group-relative policy optimization removes the need for a separate critic when a policy already samples a group per state (Shao et al., 2024; Schulman et al., 2017), heteroscedastic regression supplies our uncertainty head (Kendall & Gal, 2017; Lakshminarayanan et al., 2017), and upper confidence bounds our acquisition rule (Srinivas et al., 2009; Auer et al., 2002).

3 METHODOLOGY: EPIFORGE

3.1 PROBLEM FORMULATION

We cast epitope-conditioned design as a finite-horizon decision process in which an agent edits a candidate variable region until its predicted complex places the intended paratope against the intended epitope. The state is $s_t = (y_t, \hat{P}_t, u_t, x)$ with y_t the designable sequence, $\hat{P}_t$ the predicted complex, u_t the metric triple below, and x the context: antigen structure, epitope residues with indices, fixed framework and chain pairing, and germline assignment from standard numbering (Dunbar & Deane, 2016). Designable positions are the six CDRs, optionally restricted to CDR-H3; framework positions are never edited, letting one humanized scaffold be reused across targets.

The verifier and the quality signal. A single in-loop verifier $\mathcal{V}$ maps a candidate sequence and its context to a predicted complex and to a deterministic metric triple computed from that complex,

$$u = \mathcal{V}(y, x) = \big[\text{dockq},\ \text{epitope_recall},\ \text{h3_backbone_rmsd}\big]. \tag{1}$$

$\mathcal{V}$ is an ESMFold-based complex prediction followed by open contact, superposition and DockQ routines (Basu & Wallner, 2016), made reproducible by fixed seeds and by caching predictions on a sequence hash, and it is the only object in EPIFORGE that can change the state. For a parent and its child we form the changes $\Delta u = u_{\text{child}} - u_{\text{parent}}$, standardize each by running statistics over verified transitions, and combine them with RMSD entering negatively because smaller is better,

$$U(\Delta u) = z_{\Delta\text{DockQ}} + z_{\Delta\text{Recall}} - z_{\Delta\text{RMSD}}. \tag{2}$$

Standardizing the *changes* rather than the levels makes the signal comparable across targets of very different difficulty, the dominant source of variance here. The episode return is $\sum_t U(\Delta u_t)$ over

accepted transitions, maximized under a budget of T steps and k verifier calls per step. Because z is affine and held fixed within an episode, these per-step rewards telescope to the standardized end-to-end improvement, so no policy can accumulate return by revisiting states it has already seen (proposition 2, Appendix A.2).

Proposition 1 (Verified monotonicity)**.** *The acceptance rule of algorithm 1 advances the state only when the verified child improves the quality of the verified parent. Hence U evaluated at the accepted states is non-decreasing along an episode, and the returned design is the best* verified *state visited, not the best predicted one.*

Its proof is elementary (Appendix A.2); we state it because it pins down the division of labour. ECHO influences which candidates are verified and how credit is assigned, never what is kept, and a configuration that skips verification is not covered by it, which is how section 4.4 should be read.

Developability is measured, not optimized. Chemical liability motifs, the five interpretable descriptors (Raybould et al., 2019), a repertoire-based humanness percentile (Prihoda et al., 2022) and germline departure are computed for every returned design from the sequence and a fast antibody-only predicted structure (Kenlay et al., 2024). They do not appear in eq. (2), carry no dual prices and never gate an edit, so table 2 is a falsifiable check rather than a restatement of the objective.

3.2 INITIALIZATION BY MASKED DISCRETE DIFFUSION

EPIFORGE does not begin from a blank variable region. Stage-0 represents the complex as a chemistry-aware functional coarse-grained graph whose nodes carry residue geometry together with chemically meaningful beads (Marrink et al., 2007; Souza et al., 2021): every residue keeps three backbone beads, including a carbonyl bead placed at the midpoint of the carbonyl carbon and its oxygen so that it carries the direction the carbonyl points, while framework and antigen residues additionally carry side-chain beads for charge, aromaticity, polarity, hydrophobicity and hydrogen-bonding capacity. At designable CDR positions every side-chain-dependent bead is removed, so the model cannot read the identity it must predict. Generation is an absorbing-mask discrete diffusion over the designable residues (Austin et al., 2021; Lou et al., 2023) in which a frozen ESM-C encoder (EvolutionaryScale Team, 2024) supplies a static context computed once from the masked complex, and it is serial: the non-H3 CDRs are generated with H3 masked and an H3 specialist then generates CDR-H3 conditioned on the completed context. For affinity maturation the population is seeded from the parent's bounded mutation neighbourhood instead. The initial state is the member of that population with the best verified quality, $y_0 = \arg\max_{y \in \mathcal{Y}_0} U(\mathcal{V}(y, x))$, with $\hat{P}_0, u_0$ the artefacts of that call. Architecture, bead definitions and training are in Appendix A.5.

3.3 SKETCH: GENOTYPE PROPOSALS

SKETCH is an instruction-tuned language model acting as the policy π_ψ. Its prompt is small and fully determined by the state: the two chains, the current CDR-H3 with its indices in the heavy chain, the epitope sequence with its indices in the antigen, the metric triple u_t, and the steps remaining. It emits an action $a_t \sim \pi_\psi(\cdot \mid s_t)$ which is not free-form text but a fixed-length integer genotype

$$a_t = [\, k,\ p_1,\ v_1,\ p_2,\ v_2,\ p_3,\ v_3 \,] \in \mathbb{Z}^7, \tag{3}$$

in which k controls how many point mutations are applied and each pair (p_j, v_j) encodes a position and an amino-acid identity. This is what makes the action space closed: an integer vector cannot be malformed, so no proposal is discarded for a syntactic reason and the policy need not learn a schema in addition to a biology. Whether the integers are *meaningful* is a separate question, answered by the decoder of section 3.4 and by the verified reward rather than by a parser. The model may emit accompanying text, but a strict validator extracts seven integers and discards the rest, so every number used anywhere in EPIFORGE comes from a deterministic checker or from $\mathcal{V}$; the exact prompt and validator are in Appendix A.6.

3.4 GAUGE: SEQUENTIAL MODULAR DECODING ONTO A LEGAL MUTATION SET

GAUGE has one job, to project an arbitrary integer genotype onto a legal mutation set. It holds no learned parameters, produces no score and makes no accept or reject decision; acceptance belongs to $\mathcal{V}$ and the rollback rule alone.

Let V be the ordered list of designable positions under the active mask, $L = |V|$, and $\mathcal{B}$ the canonical alphabet of the twenty amino acids in a fixed released order. The edit size is $n = 1 + (k \bmod 3) \in \{1, 2, 3\}$, so every genotype yields at least one and at most three point mutations and the episode stays a sequence of small repairs. The pairs are then decoded left to right against shrinking sets: writing $V_1 = V$, for $j = 1, \ldots, n$,

$$\pi_j = V_j\big[\, p_j \bmod |V_j| \,\big], \qquad V_{j+1} = V_j \setminus \{\pi_j\}, \qquad \beta_j = \mathcal{A}(\pi_j)\big[\, v_j \bmod 19 \,\big], \tag{4}$$

where $\mathcal{A}(\pi_j) = \mathcal{B} \setminus \{y_t[\pi_j]\}$ is the 19-letter alphabet obtained by deleting the residue currently occupying π_j, in the induced order. Three properties follow and are the reason for this form. Taking the position modulo the number of *remaining* designable positions rather than modulo L makes repeated positions impossible rather than something to repair afterwards. Taking the identity modulo 19 against the position-specific alphabet makes identity substitutions impossible, so the policy cannot spend budget on vacuous edits and stopping is expressed through the budget and the rollback rule rather than a no-op. And the map is surjective onto all n-site non-identity mutation sets, so the projection removes no design from consideration.

The same integer p_j names different positions at different j and under different parents, so a genotype is a context-dependent code rather than a fixed address, which is why SKETCH sees the current sequence at every step; an out-of-range integer is simply a large input to a modulo. The map is deterministic given the genotype, parent sequence and active mask, and the canonical alphabet order it depends on is released with the code, so any reported edit trace can be replayed exactly; a worked example is in Appendix A.6. Each decoded mutation is additionally labelled with the typed operator whose precondition it satisfies, computed from the current predicted complex, numbering and sequence (Appendix A.6); the labels make a trajectory auditable but do not enter eq. (2), gate an edit, or train anything. The price of closure is that the policy has no way to express "do nothing", so on a target the initializer already handles the loop must spend calls or stall; section 4.2 shows this costing recovery on short loops and remark 3 discusses the alternative we rejected.

3.5 ECHO: UNCERTAINTY-AWARE WORLD MODEL

Verifying every candidate is expensive, so ECHO is a surrogate for the effect of an edit. It consumes the parent and child CDR-H3 sequences, the parent heavy chain for framework context, the epitope sequence when available and otherwise the antigen sequence, the decoded mutation list, and the parent metric triple, and returns the child triple $\hat{u}_{t+1}$ in the same fixed order with a per-entry variance. Its inputs are sequences and integers only and no structure is predicted inside it, which makes it cheap enough to score a whole slate; supplying the parent metrics rather than predicting them lets it concentrate on the effect of an edit rather than the difficulty of a target. The trunk is a transformer encoder over the embedded sequence and mutation streams and two heads read the pooled representation, one producing the child triple and one its per-entry log-variance (Appendix A.7). The predicted changes are combined by eq. (2) into $\hat{U}(s, a)$, and the scalar $\hat{\sigma}^2(s, a)$ is obtained by propagating the three per-entry variances through the fixed weights of that combination under an independence assumption, so $\hat{\sigma}$ is on the same scale as $\hat{U}$ by construction; the assumption is crude and its effect on calibration is reported rather than assumed away. Training is heteroscedastic regression against verifier-derived targets (Kendall & Gal, 2017),

$$\mathcal{L}_{\text{Echo}} = \mathbb{E}\Big[\frac{(\hat{U}_t - U_t)^2}{2\hat{\sigma}_t^2} + \tfrac{1}{2}\log\hat{\sigma}_t^2\Big] + \lambda_{\text{aux}} \sum_c \mathbb{E}\big[(\hat{u}_{t+1,c} - u_{t+1,c})^2\big], \tag{5}$$

fitted offline on verified pairs before the first rollout and refitted between cycles on every verified transition, accepted or rolled back. The variance head is the operative addition relative to a point-estimate surrogate: it makes the surrogate's own ignorance a quantity the rest of the system can act on, and section 3.6 acts on it twice with opposite signs.

3.6 SCOUT, GRPO, AND THE CLOSED LOOP

Selection is optimistic. SCOUT scores each decoded candidate by an upper confidence bound on the predicted quality change (Srinivas et al., 2009), $A(s, a) = \hat{U}(s, a) + \kappa\,\hat{\sigma}(s, a)$, and forwards the top k for verification. An edit whose effect the surrogate already knows carries little information, whereas one it cannot place is worth a call both for its own sake and because its outcome improves the

surrogate. Setting $\kappa = 0$ recovers a greedy selector, reported as an ablation. ECHO is not a Gaussian process and its variance head is not a posterior, so we claim no regret bound and treat κ as a validated hyperparameter annealed downward over training.

Credit assignment is conservative. We train π_ψ with group-relative policy optimization (Shao et al., 2024), which suits the setting because SKETCH already proposes a group per state. At state s we draw G genotypes against the same antigen and epitope, so the group baseline removes target difficulty, and score each by the surrogate with its uncertainty subtracted rather than added,

$$r_i = \hat{U}(s, a_i) - \kappa_{\text{pen}}\, \hat{\sigma}(s, a_i), \qquad \hat{A}_i = \frac{r_i - \text{mean}_j r_j}{\text{std}_j r_j + \varepsilon}, \tag{6}$$

replacing $\hat{U}$ by the realized U for the k members actually verified, and minimizing the clipped surrogate with a KL term to a frozen supervised reference that keeps the policy emitting parseable genotypes. The sign asymmetry between the two rules is the main learning-side design choice of EPIFORGE: the loop will *spend a verifier call* on an edit it cannot predict and will not *reinforce* one, so exploration is charged to the budget rather than to the policy's parameters. No developability, humanness or prior-likelihood term appears in eq. (6). An earlier version admitted designs on the strength of a learned energy with a conformal threshold; we removed it, because in a closed loop whose policy is trained against the same signal a learned gate is a second surface to exploit (remark 2).

The closed loop. One cycle runs: fit ECHO offline on the verified pairs available; update SKETCH by GRPO against eq. (6); roll out from the state pool; decode, score and select top-k; verify those k with $\mathcal{V}$; accept the best improving child or roll back; push every verified transition into the replay buffer and every accepted state into the pool; refresh ECHO; repeat. The standardization statistics of eq. (2) are refreshed only between cycles and never inside an episode, which is what keeps the telescoping argument true of the returns actually used for credit. Failed edits are as valuable as successful ones, since both are labelled transitions for eq. (5): this is the sense in which the loop learns from rejections an open-loop pipeline discards. Verifier calls per step are k regardless of how many genotypes SKETCH emits. Pseudocode is in Appendix A.9 and hyperparameters in Appendix A.12.

4 EXPERIMENTS

Setup. We build the benchmark from SAbDab (Dunbar et al., 2014), keeping complexes at 3.0 Å or better with a resolvable epitope of at least five antigen residues (Dunbar & Deane, 2016). Redundancy is controlled on three axes: CDR-H3 clustering with MMseqs2 (Steinegger & Söding, 2017) at 50% identity, antigen sequence similarity at 30% identity with at least 80% bidirectional coverage, and antigen structural redundancy by Foldseek (Van Kempen et al., 2024) retrieval with USalign (Zhang et al., 2022) re-alignment at $\min(\text{TM}_{i\to j}, \text{TM}_{j\to i}) \geq 0.50$; the three relations are combined into one graph whose connected components are assigned whole to train, validation and test in an approximate $8{:}1{:}1$ ratio (Appendix A.10). Held-out generalization uses RAbD (Adolf-Bryfogle et al., 2018) and a viral panel from CoV-AbDab (Raybould et al., 2021), both excluded by component; maturation and binding use SKEMPI 2.0, AB-Bind, an anti-HER2 deep mutational panel and a measured affinity panel (Jankauskaitė et al., 2019; Sirin et al., 2016; Mason et al., 2021; Engelhart et al., 2022).

Baselines, metrics and protocol. We compare against fixed-backbone designers (Dauparas et al., 2022; Høie et al., 2025), co-design models (Jin et al., 2021; Luo et al., 2022; Kong et al., 2023; Verma et al., 2023), preference-optimized design (Zhou et al., 2024), and the antibody-specialized diffusion pipeline that is also our initializer (Bennett et al., 2026), the most informative comparison since any improvement over it is attributable to the loop rather than the generator. We report amino-acid recovery on CDR-H3 and on all six CDRs and, under reporting predictors disjoint from the in-loop verifier (Passaro et al., 2025; Team et al., 2025), backbone deviation and epitope recall. That disjointness is the point of the protocol: a gain living only in the verifier's own coordinates would not appear in table 1, and each predictor keeps its own column because pooling would let a gain under one mask a loss under another. Developability uses liability counts, the five interpretable descriptors (Raybould et al., 2019) and a repertoire-based humanness percentile (Prihoda et al., 2022). All numbers come from a single inference pass per configuration with fixed seeds and sequence-hash caching, so the pipeline is deterministic but we report no seed variance; all evaluation is *in silico*. The ablation protocol is in Appendix A.11.

Table 1: **Epitope-conditioned design on the SAbDab structure-similarity split.** Recovery is amino-acid recovery on CDR-H3 and on all six CDRs. Backbone RMSD cells report CDR-H3 / all-CDR. Both metric families are computed under two reporting predictors disjoint from the in-loop verifier, Mean being their average. Best and second best per column are **bold** and underlined; dyMEAN and EPIFORGE tie on Protenix epitope recall (0.3766 each) and both are marked. RFantibody + ProteinMPNN is also our Stage-0 initializer, so the last row read against the first isolates the loop.

Method	Recovery (%) ↑		CDR-H3 / all-CDR backbone RMSD (Å) ↓			Epitope recall ↑		
	H3	All-CDR	Boltz-2	Protenix	Mean	Boltz-2	Protenix	Mean
Fixed-backbone inverse folding								
RFantibody + ProteinMPNN (initializer)	38.05	45.99	3.908 / 3.249	4.522 / 4.389	4.215 / 3.819	0.391	0.375	0.383
AntiFold	23.34	32.86	4.322 / 2.982	5.114 / 3.434	4.718 / 3.208	0.372	0.302	0.337
Sequence–structure co-design								
RefineGNN	26.00	35.75	4.609 / 3.592	5.948 / 3.993	5.278 / 3.792	0.393	0.287	0.340
DiffAb	38.83	54.91	3.460 / **2.222**	7.093 / 4.561	5.276 / 3.392	0.388	0.343	0.366
dyMEAN	40.08	52.48	3.542 / 2.315	4.814 / **3.015**	4.178 / **2.665**	0.387	**0.377**	0.382
AbODE	17.16	19.69	5.587 / 3.924	6.352 / 4.422	5.969 / 4.173	0.395	0.315	0.355
AbDPO	40.91	50.07	7.697 / 5.351	6.770 / 4.814	7.233 / 5.083	0.140	0.124	0.132
EPIFORGE (ours)	**45.32**	**55.11**	**3.449** / 2.528	**4.021** / 3.316	**3.735** / 2.922	**0.398**	**0.377**	**0.387**

Table 2: **Developability, humanness and recovery.** Liabilities counts CDR chemical liability motifs, TAP flags the five interpretable descriptors outside their reference range, humanness a repertoire-based percentile. None appears in eq. (2) or gates any edit; all are measured on the returned designs of the SAbDab structure-similarity split.

Method	Liabilities ↓	TAP flags ↓	Humanness ↑	CDR-H3 recovery (%) ↑
Native antibodies (reference)	5.26	1.310	29.88	100.00
RFantibody	6.564	1.310	19.552	38.05
DiffAb	5.900	1.290	39.864	38.83
RefineGNN	5.385	1.308	30.692	26.00
dyMEAN	5.370	1.333	44.487	40.08
EPIFORGE (ours)	**3.256**	**1.256**	**46.706**	**45.32**

4.1 EPITOPE-CONDITIONED DESIGN

EPIFORGE leads on both recovery columns and on CDR-H3 backbone deviation under both reporting predictors (table 1). The comparison that isolates the loop is against its own initializer: starting from exactly the RFantibody plus ProteinMPNN designs, the loop adds 7.3 points of CDR-H3 recovery, 9.1 of all-CDR recovery and 0.50 Å of CDR-H3 accuracy under Protenix, so the gain is attributable to editing rather than to the generative backbone. Two results cut against a simple reading. The epitope-recall columns are flat: our mean is 0.387 against 0.382 for dyMEAN, and under Protenix the two agree to four decimals, so the loop improves how accurately the loop is *placed* far more than how much of the epitope it *covers*, and we claim no lead there. The two measure different things, since backbone deviation is computed against the reference conformation while recall counts contacted reference residues, and a design can touch many epitope residues with a loop whose conformation is wrong: AbODE holds the second-highest Boltz-2 recall of the panel with its second-worst deviation and the lowest recovery. Separately, AbDPO reaches 40.91 recovery with by far the worst epitope recall under both predictors, the signature of optimizing a proxy that is not the interface.

4.2 LOOP LENGTH AND HELD-OUT GENERALIZATION

The length breakdown supports the hypothesis motivating an editing formulation and also bounds it (table 7, Appendix A.14). Iterative repair should help most where one forward pass must resolve several independent local defects, predicting larger gains on long loops: against dyMEAN, CDR-H3 recovery improves by 7.42 points above fifteen residues and 1.45 in the middle bin, while on twenty-three short loops it *falls* by 4.96. We read the short-loop row as the no-op cost of section 3.4 rather than noise, since a nine-residue loop offers few independent defects, the decoder cannot propose a no-op, and the budget is spent moving away from an adequate one-pass solution; with twenty-three

Table 3: **Ablation on the inference-only subset.** The first row is absolute; all others are relative changes. ESMFold is the in-loop verifier, shown for diagnosis; Boltz-2 is a reporting predictor the loop never sees; Calls is the mean verifier calls per target. Values are not comparable with table 1.

	Sequence	Boltz-2 (reporting)		ESMFold (in-loop)		
Configuration	H3 rec.	DockQ↑	Epi. recall↑	DockQ↑	Epi. recall↑	Calls
Initializer only (absolute)	0.5026	0.4792	0.4792	0.1358	0.1977	0
Effect of edit depth, full EPIFORGE						
depth 1	−0.0143	−0.0002	+0.0029	−0.0032	+0.0007	0.37
depth 2	−0.0278	−0.0070	+0.0100	−0.0038	+0.0018	0.71
depth 3 (full)	−0.0334	+0.0121	+0.0012	−0.0038	+0.0018	0.97
Module ablations at depth 1						
without SKETCH (random legal genotype)	−0.0026	+0.0005	+0.0027	−0.0008	0.0000	0.05
ECHO-only guidance ($\kappa = 0$, no verification)	−0.0062	+0.0261	+0.0015	−0.0015	0.0000	0.15

targets we claim only the direction. Backbone accuracy improves in every bin, and both held-out panels, RAbD and a viral CoV-AbDab panel excluded by component, gain about nine recovery points at lower deviation, evidence that the gain is not an artefact of our split. The per-family breakdown is in Appendix A.14.

4.3 DEVELOPABILITY AND HUMANNESS

Table 2 is the check the loop is allowed to fail: nothing in eq. (2) rewards a humanized or liability-free sequence, and an interface-directed agent may drift from repertoire statistics in pursuit of a contact. It does not: EPIFORGE attains the best humanness of any method, 16.8 percentile points above the native reference and 2.2 above the best baseline, carrying 3.26 liability motifs against the native 5.26, at the highest recovery in the table. We attribute this to the initializer and to the small bounded edits the decoder permits rather than to any pressure in the objective, and regard it as a property to monitor rather than one that is guaranteed. Two cautions govern how the first three columns should be read. They are one-sided: a low liability count and a clean descriptor profile are also what a conservative, low-complexity CDR produces, so neither is evidence of a usable design on its own, and both are informative only alongside the recovery column, which is why we report them together. And none of them is measured affinity, expression or stability; they are descriptors of a sequence and of a fast antibody-only predicted structure, reported here because the objective ignores them, not because they establish developability.

4.4 ABLATION AND THE RECOVERY TO INTERFACE DISSOCIATION

Table 3 contains the result we consider most important to report plainly, because it complicates the headline. On that subset the editing loop *reduces* CDR-H3 recovery monotonically with depth, by 1.4, 2.8 and 3.3 points, while Boltz-2 DockQ improves only at depth three and epitope recall improves at every depth. Recovery and interface quality therefore dissociate under editing, in the direction the objective predicts: eq. (2) rewards a verified change in interface quality and nothing else, in particular not similarity to the native sequence, so an edit that leaves the native residue while improving the verified interface is what the agent is trained to make, and the decoder never conserves a native residue by default. This does not contradict table 1, where the comparison is across methods on a different subset, but recovery is not the quantity the loop optimizes. Depth is not free either, since calls per target rise from 0.37 to 0.97 while the DockQ gain appears only at the last depth. The module rows are not budget-matched, and the largest Boltz-2 DockQ change belongs to ECHO-only guidance at a fifth of the budget; we do not read that as evidence that SKETCH is unnecessary, since that configuration verifies almost nothing, proposition 1 does not apply to it, and its in-loop column moves the other way. A budget-matched comparison is the main gap here.

4.5 CASE STUDY: JOINT RECOVERY OF POSE, LOOP AND SEQUENCE

Table 3 separates sequence recovery, loop conformation and interface placement in aggregate; a single complex shows what that separation looks like structurally, and what it looks like when it is absent. On 1NCD (fig. 2), a held-out target excluded by component, the three families agree: the designed Fv lands on the native Fv rather than merely somewhere on the antigen, at DockQ

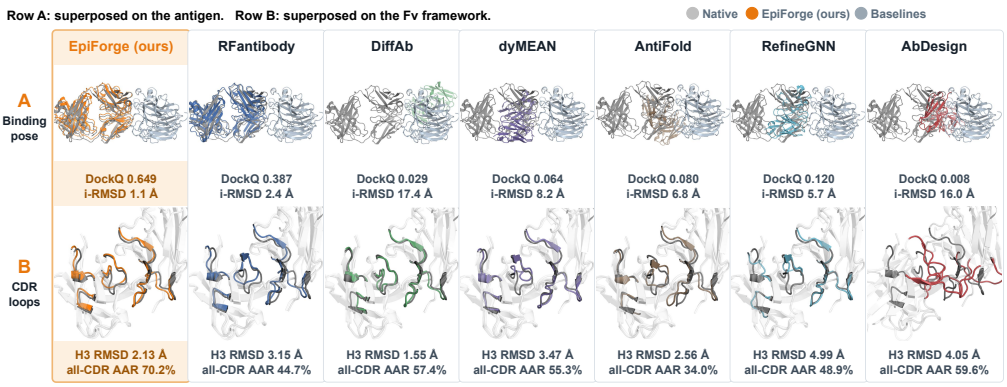

Figure 2: **Case study 1NCD: pose and loops recovered together.** All seven methods on one held-out target, with the native structure in grey. Row A superposes each design on the antigen and reports DockQ and interface RMSD; row B superposes on the Fv framework and reports CDR-H3 backbone deviation and all-CDR recovery. EPIFORGE reaches DockQ 0.649 at interface RMSD 1.1 Å, against 0.387 for the strongest baseline, with the highest all-CDR recovery of the panel at 70.2% and every contact residue of the native epitope re-engaged. This is the best case rather than the typical one; section A.4 gives two targets where the three quantities come apart.

0.649 and interface RMSD 1.1 Å against 0.387 for the strongest baseline, while all-CDR recovery is simultaneously the highest of the panel at 70.2% and every contact residue of the native epitope is re-engaged. The comparison is informative precisely because the initializer is in the same panel: the pose it hands the loop is refined rather than replaced, and the improvement is a small number of verified point mutations rather than a different backbone. We present this target as the best case and not the typical one, and report two others in Appendix A.4 where the agreement breaks in each of the two available directions. On 1A14 our CDR conformations are the panel's most accurate, 1.18 Å on CDR-H3 against about 3.39 Å for the next best method, while RFantibody wins the rigid-body pose outright; this is possible because backbone deviation is measured after superposition on the Fv framework whereas DockQ is measured against the antigen, so the two can move apart. On 6CMG the sequence is recovered best in the panel, 64.0% on CDR-H3, while a long CDR-H3 defeats every method geometrically and no design exceeds DockQ 0.05; the honest reading there is that no method solves the target. Taken together the three complexes are why we report the metric families side by side rather than collapsing them into one score.

4.6 AFFINITY MATURATION

Affinity maturation is the setting an editing formulation suits best, and it is the one setting in which EPIFORGE is the only method to return an improved design at all: under the protocol of Appendix A.13 it improves a predicted interaction energy where every co-design baseline improves none, and it is best or second best on every energetic column. We do not read this as settled. The improved fraction is small in absolute terms, the quantities are predicted rather than measured, and the protocol's bounded mutation budget is one an editing formulation is built for and a one-pass generator is not, so part of the margin reports fit to the protocol rather than maturation ability. Appendix A.3 summarizes the main limitations: single-verifier optimization, entirely *in silico* evaluation, single-run reporting and non-budget-matched ablations.

5 CONCLUSION

EPIFORGE casts epitope-conditioned antibody design as closed-loop verified repair: every action is legal, only structural verification advances the state, and uncertainty guides verification and credit. It improves over its initializer by 7.3 points in CDR-H3 recovery and 9.1 in all-CDR recovery, with stronger backbone accuracy under unseen predictors and the best humanness among compared methods. Its main limitation is the verifier: errors shared by $\mathcal{V}$ remain invisible during search. Future work should use verifier ensembles, explicit call pricing and wet-laboratory validation.

AI USE STATEMENT

The authors used a general-purpose large language model assistant only for language editing of the manuscript: grammar, concision and consistency of terminology. All text was reviewed and verified by the authors. Separately, and as a matter of the method rather than of authorship, an instruction-tuned language model is a component of EPIFORGE itself: SKETCH (section 3.3) is the policy π_ψ that emits the integer genotype of eq. (3). Its outputs are read only through a strict schema validator, so no free-form text it produces enters the pipeline and no reported number depends on anything it asserts.

ETHICS STATEMENT

The authors have read the ICLR Code of Ethics and this work adheres to it. The study involves no human subjects and no personally identifying information. Every dataset used is a public, previously released resource — SAbDab (Dunbar et al., 2014), RAbD (Adolf-Bryfogle et al., 2018), CoV-AbDab (Raybould et al., 2021), SKEMPI 2.0 (Jankauskaitė et al., 2019), AB-Bind (Sirin et al., 2016), an anti-HER2 deep mutational panel (Mason et al., 2021) and a measured affinity panel (Engelhart et al., 2022) — used within the terms of its licence, with all preprocessing described in Appendix A.10. On dual use, EPIFORGE is an editing layer built on publicly released models and openly available tooling, it targets antibodies, and it is evaluated entirely *in silico*: no design reported here has been synthesized, expressed or assayed, and we release no sequences beyond what is already deposited in the public databases cited above. All results are predictions from structure and sequence models, and we claim no experimentally validated binding, affinity or developability. The authors declare no competing interests.

REPRODUCIBILITY STATEMENT

Anonymized source code, covering the Stage-0 initializer, the four loop components, the verifier wrapper and the evaluation harness, is available at `https://anonymous.4open.science/r/EpiForge-82C6/`. Dataset construction, the three redundancy relations, the component-wise split and the filtering rules are given in section 4 and Appendix A.10; the action decoder is deterministic given a genotype and a parent, and the canonical alphabet order it depends on is released explicitly (Appendix A.6), so any reported edit trace can be replayed exactly. ECHO's interface, architecture and training objective are in section 3.5 and Appendix A.7, inference and training pseudocode in Appendix A.9, the ablation protocol in Appendix A.11, and hyperparameter groups, library versions and container digests in Appendix A.12. The verifier is made deterministic by fixed seeds and prediction caching, which is assumption (A3) of Appendix A.2; the assumptions and complete proofs of propositions 1 and 2 are given there. No licensed molecular-modelling suite is required at any point.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Jared Adolf-Bryfogle, Oleks Kalyuzhniy, Michael Kubitz, Brian D Weitzner, Xiaozhen Hu, Yumiko Adachi, William R Schief, and Roland L Dunbrack Jr. Rosettaantibodydesign (rabd): A general framework for computational antibody design. *PLoS computational biology*, 14(4):e1006112, 2018.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 11(8):e0161879, 2016.

Nathaniel R Bennett, Joseph L Watson, Robert J Ragotte, Andrew J Borst, DéJenaé L See, Connor Weidle, Riti Biswas, Yutong Yu, Ellen L Shrock, Russell Ault, et al. Atomically accurate de novo design of antibodies with rfdiffusion. *Nature*, 649(8095):183–193, 2026.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

James Dunbar and Charlotte M Deane. Anarci: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, 2016.

James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M Deane. Sabdab: the structural antibody database. *Nucleic acids research*, 42 (D1):D1140–D1146, 2014.

Emily Engelhart, Ryan Emerson, Leslie Shing, Chelsea Lennartz, Daniel Guion, Mary Kelley, Charles Lin, Randolph Lopez, David Younger, and Matthew E Walsh. A dataset comprised of binding interactions for 104,972 antibodies against a sars-cov-2 peptide. *Scientific Data*, 9(1):653, 2022.

Richard Evans, Michael O'neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

EvolutionaryScale Team. ESM Cambrian: Revealing the mysteries of proteins with unsupervised learning, 2024.

Magnus Haraldson Høie, Alissa M Hummer, Tobias H Olsen, Broncio Aguilar-Sanjuan, Morten Nielsen, and Charlotte M Deane. Antifold: improved structure-based antibody design using inverse folding. *Bioinformatics Advances*, 5(1):vbae202, 2025.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *International conference on machine learning*, pp. 8946–8970. PMLR, 2022.

John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H Moal. Skempi 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019.

Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. *arXiv preprint arXiv:2110.04624*, 2021.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Antibody-antigen docking and design via hierarchical structure refinement. In *International Conference on Machine Learning*, pp. 10217–10227. PMLR, 2022.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

Henry Kenlay, Frédéric A Dreyer, Daniel Cutting, Daniel Nissley, and Charlotte M Deane. Abodybuilder3: improved and scalable antibody structure predictions. *Bioinformatics*, 40(10):btae576, 2024.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation. *arXiv preprint arXiv:2208.06073*, 2022.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. *arXiv preprint arXiv:2302.00203*, 2023.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Marc F Lensink, Raúl Méndez, and Shoshana J Wodak. Docking and scoring protein complexes: Capri 3rd edition. *Proteins: Structure, Function, and Bioinformatics*, 69(4):704–718, 2007.

Haitao Lin, Lirong Wu, Yufei Huang, Yunfan Liu, Odin Zhang, Yuanqing Zhou, Rui Sun, and Stan Z Li. Geoab: towards realistic antibody design and reliable affinity maturation. *bioRxiv*, pp. 2024–05, 2024.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. *arXiv preprint arXiv:2310.16834*, 2023.

Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35:9754–9767, 2022.

Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

Derek M Mason, Simon Friedensohn, Cédric R Weber, Christian Jordi, Bastian Wagner, Simon M Meng, Roy A Ehling, Lucia Bonati, Jan Dahinden, Pablo Gainza, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature biomedical engineering*, 5(6):600–612, 2021.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature methods*, 19(6):679–682, 2022.

Tobias H Olsen, Fergus Boyles, and Charlotte M Deane. Observed antibody space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science*, 31 (1):141–146, 2022.

Tobias H Olsen, Iain H Moal, and Charlotte M Deane. Addressing the antibody germline bias and its effect on language models for improved antibody design. *Bioinformatics*, 40(11):btae618, 2024.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *BioRxiv*, 2025.

David Prihoda, Jad Maamary, Andrew Waight, Veronica Juan, Laurence Fayadat-Dilman, Daniel Svozil, and Danny A Bitton. Biophi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. In *MAbs*, volume 14, pp. 2020203. Taylor & Francis, 2022.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Matthew IJ Raybould, Claire Marks, Konrad Krawczyk, Bruck Taddese, Jaroslaw Nowak, Alan P Lewis, Alexander Bujotzek, Jiye Shi, and Charlotte M Deane. Five computational developability guidelines for therapeutic antibody profiling. *Proceedings of the National Academy of Sciences*, 116(10):4025–4030, 2019.

Matthew IJ Raybould, Aleksandr Kovaltsuk, Claire Marks, and Charlotte M Deane. Cov-abdab: the coronavirus antibody database. *Bioinformatics*, 37(5):734–735, 2021.

Jeffrey A Ruffolo, Jeffrey J Gray, and Jeremias Sulam. Deciphering antibody affinity maturation with language models and weakly supervised learning. *arXiv preprint arXiv:2112.07782*, 2021.

Jeffrey A Ruffolo, Lee-Shin Chu, Sai Pooja Mahajan, and Jeffrey J Gray. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature communications*, 14(1):2389, 2023.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Richard W Shuai, Jeffrey A Ruffolo, and Jeffrey J Gray. Iglm: Infilling language modeling for antibody sequence design. *Cell systems*, 14(11):979–989, 2023.

Sarah Sirin, James R Apgar, Eric M Bennett, and Amy E Keating. Ab-bind: antibody binding mutational database for computational affinity predictions. *Protein Science*, 25(2):393–409, 2016.

Paulo CT Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart MH Bruininks, Tsjerk A Wassenaar, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods*, 18(4):382–388, 2021.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

Martin Steinegger and Johannes Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11):1026–1028, 2017.

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, et al. Protenix-advancing structure prediction through a comprehensive alphafold3 reproduction. *BioRxiv*, pp. 2025–01, 2025.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*, pp. 2024–10, 2024.

Michel Van Kempen, Stephanie S Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron LM Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with foldseek. *Nature biotechnology*, 42(2):243–246, 2024.

Yogesh Verma, Markus Heinonen, and Vikas Garg. Abode: Ab initio antibody design using conjoined odes. In *International Conference on Machine Learning*, pp. 35037–35050. PMLR, 2023.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Chengxin Zhang, Morgan Shine, Anna Marie Pyle, and Yang Zhang. Us-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature methods*, 19(9): 1109–1115, 2022.

Xiangxin Zhou, Dongyu Xue, Ruizhe Chen, Zaixiang Zheng, Liang Wang, and Quanquan Gu. Antigen-specific antibody design via direct energy-based preference optimization. *Advances in Neural Information Processing Systems*, 37:120861–120891, 2024.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Forty-first International Conference on Machine Learning*, 2024.

A Appendix

A.1 Notation

We collect the notation used by the decision process, decoder and verifier in table 4. Symbols with a time index refer to the current editing episode, and quantities with hats are predictions rather than verified measurements unless stated otherwise. The verifier output u_t is always the metric triple computed after a structure prediction, whereas $\hat{U}$ and $\hat{\sigma}$ are the sequence-level surrogate's estimate and uncertainty used only for ranking and credit.

Table 4: Notation.

Symbol	Meaning
A, E, F, g	antigen, epitope residue set, framework, germline assignment
x	context, comprising A, E with indices, F, g, optional binding target
V, V_j, L	designable positions, the subset still unselected at decode step j, and $L = \lvert V \rvert$
y_t, $\hat{P}_t$, u_t, s_t	designable sequence, predicted complex, metric triple, state $(y_t, \hat{P}_t, u_t, x)$
a_t	integer genotype of eq. (3)
n, π_j, β_j	decoded mutation count, position, and amino acid of eq. (4)
$\mathcal{B}$, $\mathcal{A}(\pi)$	canonical twenty-letter alphabet; the nineteen letters excluding the residue at π
$\mathcal{V}$	in-loop verifier: structure prediction plus deterministic interface metrics
U, $\hat{U}$, $\hat{\sigma}$	verified quality change eq. (2), its ECHO estimate, and its standard deviation
κ, κ_{pen}	optimism coefficient in selection, pessimism coefficient in GRPO credit
k, G, T	verification width, group size, step budget
$\mathrm{B_{CO}}$	carbonyl backbone bead of eq. (7)

A.2 Proofs and scope of the two statements

Proposition 2 (The per-step signal decomposes the final gain)**.** *Let $\mathcal{A}$ index the accepted transitions of an episode and let the standardization map $z(\cdot) = (\cdot - \mu)/\sigma$ be held fixed within the episode. Then $\sum_{t \in \mathcal{A}} U(\Delta u_t)$ equals the standardized end-to-end improvement $\bar{z}(u_T - u_0)$ up to a constant proportional to $\lvert\mathcal{A}\rvert$, which vanishes for centred statistics. No policy can therefore accumulate return by cycling through states it has already visited.*

Assumptions. (A1) The horizon is finite and episodes terminate. (A2) Within an episode the standardization map $z(\cdot) = (\cdot - \mu)/\sigma$ applied in eq. (2) is held fixed, with running statistics refreshed only between cycles. (A3) The verifier $\mathcal{V}$ is deterministic given a sequence and a context, which we enforce by fixing seeds and caching predictions by sequence hash, so a state revisited within an episode receives the identical triple.

Proof of proposition 2. Write $U(\Delta u) = \langle w, z(\Delta u) \rangle$ with $w = (1, 1, -1)$ and z affine and fixed by (A2). Affinity gives $z(u_{t+1} - u_t) = z(u_{t+1}) - z(u_t) + c$ where $c = (\mu_1/\sigma_1, \mu_2/\sigma_2, \mu_3/\sigma_3)$ is constant. Summing over accepted transitions, the differences telescope to $\langle w, z(u_T) - z(u_0) \rangle$ and the residual is $\lvert\mathcal{A}\rvert \langle w, c \rangle$, which is a constant of the episode length. With centred statistics $c = 0$ and the residual vanishes. Because the sum depends on the endpoints alone, a trajectory that leaves and returns to a state accrues zero net return over that excursion under (A3). □

Proof of proposition 1. Immediate from the acceptance rule: line 12 of algorithm 1 replaces s_t by the child maximizing the verified quality only when that maximum exceeds the parent's, and otherwise rolls back. Hence the sequence of accepted states has non-decreasing verified quality, and the returned state is the maximizer over the visited verified set. □

Remark 1 (What verification does and does not guarantee)**.** *Proposition 1 says the returned design is the best under $\mathcal{V}$ among those actually predicted. It says nothing about any other predictor, and nothing about reality. The gap between the in-loop and reporting columns of table 3 is the empirical size of the first of these, and we report it rather than arguing it away. A single verifier is the main structural limitation of the present system: an error shared by $\mathcal{V}$ across a whole target is invisible to the loop, and the only defence we currently have is that the reporting panel is disjoint and would reveal it in aggregate.*

Remark 2 (Why acceptance is never learned)**.** *An earlier version of* EPIFORGE *carried a learned energy with a conformal acceptance threshold, so a design could be admitted on the strength of a learned score. In a closed loop whose policy is trained against the same signal, such a gate is a second surface to exploit, and its distribution-free guarantee is voided exactly where it is most wanted: on a novel epitope the calibration set and the deployed designs are not exchangeable. Restricting the learned components to* ranking *removes that failure mode at the cost of more verifier calls, which the optimism rule of section 3.6 manages by spending them where the surrogate is least certain.*

Remark 3 (The cost of a decoder that cannot do nothing)**.** *Because eq.* (4) *cannot emit an identity substitution, stopping is expressed only by exhausting the budget or by having every candidate rolled back. On targets where the initializer is already adequate this costs verifier calls and, as table 7 shows on short loops, can cost recovery. The alternative, retaining identity substitutions as explicit no-ops, wastes one of twenty alphabet symbols and in our runs was frequently emitted early in training as a degenerate low-variance action. We preferred the present form and report the consequence rather than hiding it.*

Remark 4 (Scope of proposition 2)**.** *The statement is about the reward the policy is credited with, not about the search. It says that an episode's return is a function of its endpoints, so no cycle can be farmed for return; it does not say the loop finds a good endpoint, and it says nothing once the standardization statistics are refreshed between cycles, which is precisely why (A2) is confined to a single episode.*

A.3 LIMITATIONS

Four limitations are structural. First, the loop optimizes one verifier, so an error $\mathcal{V}$ makes consistently on a target is invisible to the agent; the disjoint reporting panel can expose such errors only after the fact, not prevent them during search. Second, all evaluation is *in silico*: no design reported here has been synthesized, expressed or assayed, and predicted interface quality is not measured affinity. Third, numbers are single-run, so we do not quantify seed variance or separate stochastic variation from method differences. Fourth, the module ablations are not budget-matched, making them diagnostic rather than controlled comparisons. These limitations do not affect the deterministic decoding or verified-rollback claims, but they bound the biological and comparative conclusions that can be drawn from the experiments.

A.4 ADDITIONAL CASE STUDIES

Figure 2 in the main text is the case where sequence recovery, loop conformation and interface placement improve together. The two targets below are the cases where they do not, one in each available direction, and they use the same panel layout.

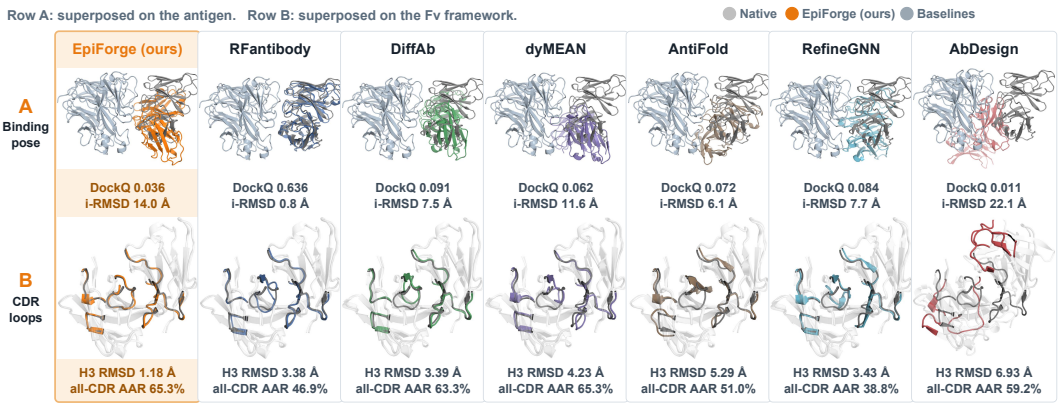

Figure 3: **Case study 1A14: loop geometry without the pose.** Panel layout as in fig. 2. EPIFORGE gives the most accurate CDR conformations of the panel, backbone deviation 1.18 Å on CDR-H3 against about 3.39 Å for the next best method, and the only sub-Ångström all-CDR value at 0.78 Å, while RFantibody wins the rigid-body pose outright with DockQ 0.636 against our 0.036. We report the target as it is: the advantage here is conformation, not placement.

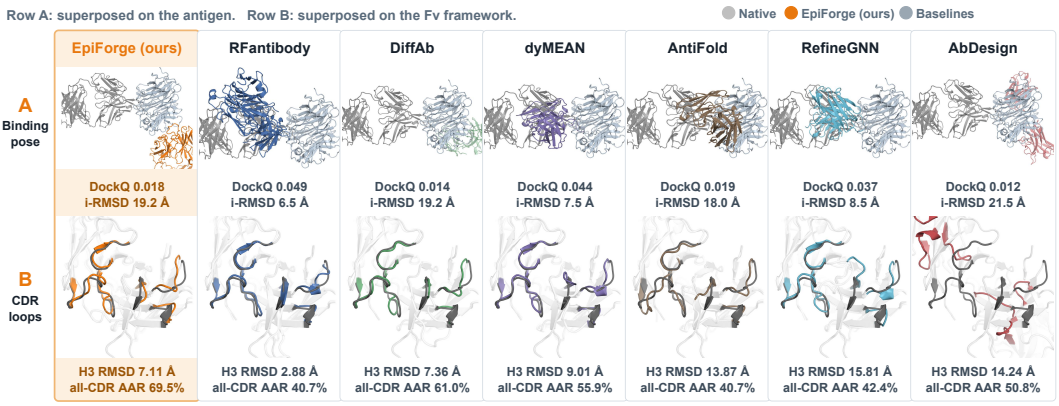

Figure 4: **Case study 6CMG: sequence without geometry.** Panel layout as in fig. 2. EPIFORGE attains the highest recovery of the panel, 64.0% on CDR-H3 against 48.0% for dyMEAN and 44.0% for DiffAb, and 69.5% over all CDRs, but its CDR-H3 backbone deviation of 7.11 Å sits mid-panel and no method exceeds DockQ 0.05 on this target. 6CMG is therefore a recovery case rather than a docking case.

On 1A14 (fig. 3) the loops are almost exactly right while the pose is not, which is possible because backbone deviation is measured after superposition on the Fv framework whereas DockQ is measured against the antigen, so the two can be optimized apart; a verifier that scores the interface will not by itself fix a rigid-body placement its own prediction already got wrong. On 6CMG (fig. 4) the sequence is recovered best in the panel while a long CDR-H3 defeats every method geometrically, and the honest reading is that no method solves the target: a recovery number on such a target certifies nothing about binding. Taken with fig. 2, the three complexes show on individual structures the same separation between sequence recovery, loop conformation and interface placement that table 3 shows in aggregate, and are why we report the three families side by side instead of collapsing them into one score.

A.5 STAGE-0 ARCHITECTURE AND TRAINING

Stage-0 receives the antigen and framework structural context, the chemistry-aware functional coarse graph, and the corrupted CDR sequence state, and follows a dual-stream design in which a geometric graph encoder processes structure and chemistry while an ESM-C conditioned residue encoder processes sequence.

Backbone beads. For every residue three backbone beads are retained, $\{\mathrm{N}, \mathrm{C}_\alpha, \mathrm{B}_{\mathrm{CO}}\}$. The first two are the amide nitrogen and the α carbon at their deposited coordinates. The third is not the carbonyl carbon but the carbonyl bead

$$\mathrm{B}_{\mathrm{CO}} = \tfrac{1}{2}\big(p_{\mathrm{C}} + p_{\mathrm{O}}\big), \qquad \text{falling back to} \qquad \mathrm{B}_{\mathrm{CO}} = p_{\mathrm{C}} \ \text{ when the carbonyl oxygen is unresolved,} \tag{7}$$

so that a single bead encodes the direction in which the peptide carbonyl points and therefore carries backbone hydrogen-bonding geometry that p_{C} alone does not. When the carbonyl oxygen is absent from the structure, for instance at a chain terminus or where the deposited entry leaves it unresolved, the fallback is recorded as a binary node feature so the encoder can distinguish a genuine midpoint from a substituted one. Terminal carboxylate oxygens are not treated as carbonyl oxygens.

Framework and antigen residues additionally receive side-chain functional beads, each node being represented as $v_i = [e_i^{\mathrm{type}}, e_i^{\mathrm{res}}, e_i^{\mathrm{chain}}, p_i, c_i]$ with bead identity, residue information, coordinate, and chemical descriptors. At the designable positions $\{\mathrm{H1}, \mathrm{H2}, \mathrm{H3}, \mathrm{L1}, \mathrm{L2}, \mathrm{L3}\}$ only the three backbone beads survive, so native identities are inaccessible where they are to be predicted. Edges are $E = E_{\mathrm{spatial}} \cup E_{\mathrm{residue}} \cup E_{\mathrm{sequence}} \cup E_{\mathrm{interface}}$, each carrying distance encodings and relation embeddings so that geometric proximity is distinguishable from chemical interaction.

The graph stream updates beads by message passing, $h_i^{l+1} = \mathrm{GraphBlock}(h_i^l, \{h_j^l, e_{ij}\}_{j \in \mathcal{N}(i)})$. The sequence stream represents the corrupted state $Y_t \in \{20\ \mathrm{AA}, \mathrm{MASK}\}^{|V|}$ by amino-acid embeddings

plus a sinusoidal diffusion-time embedding. Because the antigen and framework context does not change across reverse steps, ESM-C is evaluated once on the masked complex and its output is treated as a static context. The two streams are aligned by residue to bead attention pooling, $r_i^{\text{graph}} = \text{AttnPool}(\{b_j : j \in \text{residue}(i)\})$ and $h_i = \text{Fuse}(h_i^{\text{seq}}, r_i^{\text{graph}}, H_i^{\text{ESM}})$, with a gated reverse update $b_j^{l+1} = b_j^l + \sigma(W_g[h_j^{\text{res}}, b_j^l]) \odot W_u h_j^{\text{res}}$ injecting sequence information back into the beads.

Training uses an absorbing-mask process: for a clean sequence Y_0 a step $t \sim \mathcal{U}(1, T)$ is drawn, each designable residue is independently masked with probability t/T, and the denoiser is trained with $\mathcal{L} = -\sum_{i \in M_t} \log p_\theta(Y_{0,i} \mid Y_t, G, H^{\text{ESM}}, t)$ over masked positions only. Generation is serial: the non-H3 CDRs are produced first with H3 masked, the result is written into the context, and an H3 specialist then produces CDR-H3 conditioned on antigen, framework, and the generated non-H3 loops, giving $Y_0 = (Y_{\text{non-H3}}, Y_{\text{H3}})$.

A.6 Decoding, operator labels, and the prompt

Decoding. Equation (4) is applied left to right over the genotype and is fully deterministic given (a_t, y_t, V). The count comes first, $n = 1 + (k \bmod 3)$, so a genotype always specifies at least one mutation. At step j the position index is taken modulo $|V_j| = L - j + 1$, the number of designable positions not yet chosen in this step, against the list V_j in ascending numbering order; the chosen position is then removed, so positions are distinct by construction and the count n is always realized. The identity index is taken modulo 19 against $\mathcal{A}(\pi_j) = \mathcal{B} \setminus \{y_t[\pi_j]\}$, the canonical alphabet with the residue currently at π_j deleted and the remaining letters kept in their canonical relative order, so the substitution is always a real change. The canonical alphabet order and the list construction are released with the code, so a genotype together with a parent sequence reproduces a mutation set exactly. As a worked example, $[17, 42, 99, 7, 138, 201, 53]$ with $L = 14$ gives $n = 1 + (17 \bmod 3) = 3$, acting at $V_1[42 \bmod 14]$, $V_2[7 \bmod 13]$ and $V_3[201 \bmod 12]$ with identities $\mathcal{A}(\pi_j)[v_j \bmod 19]$; an out-of-range integer is not an error but simply a large input to a modulo.

Two consequences are worth stating because they differ from a naive $(\bmod L, \bmod 20)$ decoder. First, the same integer p_j denotes different positions at different j and for different parents, so the mapping from genotype to mutation set is context-dependent; SKETCH therefore receives the current sequence and indices in every prompt, and the decoded mutation set is not carried across steps as a fixed code. Second, the decoder is surjective but many to one onto n-site non-identity mutation sets, so distinct genotypes can name the same edit; the deduplication step of Appendix A.8 removes such duplicates from a slate before any verifier call is spent.

Operator labels. Each decoded mutation is labelled with the typed operator whose precondition it satisfies, computed from the current predicted complex, the numbering, and the sequence. Interface-forming labels include installing an aromatic contact against an apolar epitope patch, forming a salt bridge to a charged epitope residue, forming a hydrogen bond to a named polar epitope atom, and extending a hydrophobic patch. Interface-repairing labels include relieving a clash, repairing a framework-adjacent anchor, retuning a loop apex, and stabilizing the heavy to light interface. Developability labels include removing a liability motif, reducing a surface hydrophobic patch, adjusting regional charge, and reverting to germline. A mutation matching no precondition is labelled generic. All checks use open tooling: contacts are heavy-atom distance cutoffs, accessibility comes from a standard surface routine on the fast antibody-only prediction, liability motifs are regular expressions over CDR sequences, and germline identity comes from the numbering assignment. The labels are recorded for auditing and reported in aggregate in Appendix A.14; they do not enter the objective and do not gate anything.

The prompt. The prompt has a static portion that is byte-identical across steps for a given target, which allows provider-side prefix caching, and a dynamic portion carrying: the heavy chain sequence, the light chain sequence, the current CDR-H3 sequence with its start and end indices in the heavy chain, the epitope sequence with its residue indices in the antigen, the current metric triple u_t, and the number of remaining steps. The response is parsed by a strict schema validator that extracts seven integers and discards all other fields, so a model that emits a claimed improvement, an asserted interface effect, or an out-of-range index cannot have any of it enter the pipeline.

A.7 ECHO INPUTS, ARCHITECTURE, AND TRAINING

Table 5 lists the interface. ECHO is deliberately a *sequence-level* model: it consumes no coordinates and predicts no structure, which is what makes it cheap enough to score an entire slate of decoded candidates for a small fraction of the cost of one verifier call. The parent and child CDR-H3 sequences, the parent heavy chain, and the epitope sequence (or, when no epitope sequence is available, the antigen sequence) are tokenized and embedded; the decoded mutation list is encoded as position embeddings combined with source and target identity embeddings and injected both as a global token and as an additive feature on the tokens of the mutated positions, so the network can distinguish the same substitution at two different sites; and the parent metric triple is broadcast across the sequence. The trunk is a transformer encoder over the concatenated streams and two heads read the pooled representation, one producing the child metric triple and one its per-entry log-variance, with the variance head initialized large so that early training is dominated by the mean term.

ECHO is fitted offline before the first rollout on parent–child pairs verified during data preparation, and refreshed after every cycle on the transitions accumulated in the replay buffer, which include rolled-back edits: a rejected child is a fully labelled training pair and is retained. The scalar $\hat{\sigma}$ used by SCOUT and by eq. (6) is obtained by propagating the three per-entry variances through the fixed weights $(1, 1, -1)$ of eq. (2) and the standardization scales, treating the entries as independent; the independence assumption is crude and its effect on calibration is reported in Appendix A.14.

Table 5: ECHO interface. Every input is a sequence, an integer or a scalar; no structure enters the surrogate. The output order is fixed and is the order used everywhere in the code.

Field	Meaning
`parent_h3`	CDR-H3 sequence before the edit
`child_h3`	CDR-H3 sequence after applying the decoded mutation set
`parent_heavy`	heavy chain before the edit, supplying framework context
`epitope_sequence`	sequence of the epitope residues, used as antigen context when available
`antigen_sequence`	antigen sequence, used when the epitope sequence is unavailable
`parent_metrics`	parent triple: dockq, epitope_recall, h3_backbone_rmsd
`mutations`	decoded list, each entry a position with source and target residue
output	`["dockq", "epitope_recall", "h3_backbone_rmsd"]` and per-entry variance

A.8 SEARCH AND SELECTION

The slate at each step is built by sampling G genotypes from SKETCH, decoding each by eq. (4), deduplicating by the resulting child sequence, and capping proposals per position so that the slate cannot degenerate into G variations at one site. Candidates are ranked by the upper confidence bound of section 3.6 and ties are broken by lower predicted variance, so that between two candidates of equal optimistic value the better-understood one is verified first. The optimism coefficient κ is annealed downward over training, so that early cycles buy surrogate coverage and later cycles exploit it. Verification is the only expensive step and its cost per step is k calls to $\mathcal{V}$, independent of G; predictions are cached by sequence hash across configurations and across steps, so a child reached twice is predicted once.

A.9 ALGORITHMS

Two details of algorithm 2 matter in practice. The replay buffer is fed by *all* verified children, not only accepted ones, so the surrogate is trained on the rejection boundary rather than only on successes, and the fraction of rolled-back edits is therefore a training resource rather than wasted budget. And the standardization statistics of eq. (2) are refreshed only between cycles, never within an episode, which is assumption (A2) of Appendix A.2 and is what keeps proposition 2 true of the returns actually used for credit assignment.

A.10 DATASET CONSTRUCTION AND PREPROCESSING

Complexes are numbered with a standard scheme and chains are paired by the deposited assembly. The epitope specification is the set of antigen residues within a heavy-atom cutoff of any antibody

Algorithm 1 EPIFORGE inference: one episode

Require: context x, step budget T, width k, group size G, optimism κ, stall tolerance ϱ

$\mathcal{Y}_0 \leftarrow$ Stage-0 population for (A, E, F); $u(y) \leftarrow \mathcal{V}(y, x)$ for $y \in \mathcal{Y}_0$; $y_0 \leftarrow \arg\max_y U(u(y))$
$s_0 \leftarrow (y_0, \hat{P}_0, u_0, x)$; stall $\leftarrow 0$
for $t = 0$ **to** $T - 1$ **do**
  sample $\{a_i\}_{i=1}^{G} \sim \pi_\psi(\cdot \mid s_t)$ ▷ seven integers each; non-integer output is resampled
  $\mathcal{D}_t \leftarrow \{\text{DECODE}(a_i, y_t, V)\}$ by $n = 1 + (k \bmod 3)$ and eq. (4); deduplicate by resulting sequence
  for all $d \in \mathcal{D}_t$ **do**
    $y' \leftarrow \text{APPLY}(y_t, d)$; label each mutation with its operator; $(\hat{u}, \hat{\sigma}) \leftarrow f_\theta(s_t, d)$
    $A(s_t, d) \leftarrow \hat{U}(s_t, d) + \kappa\hat{\sigma}(s_t, d)$ ▷ optimistic, section 3.6
  end for
  $\mathcal{K} \leftarrow$ top-k of $\mathcal{D}_t$ by A
  for all $d \in \mathcal{K}$ **do** ▷ the only expensive step
    $u'_d \leftarrow \mathcal{V}(\text{APPLY}(y_t, d), x)$; $U'_d \leftarrow U(u'_d - u_t)$ by eq. (2); push (s_t, d, u'_d) to $\mathcal{B}$
  end for
  if $\max_{d \in \mathcal{K}} U'_d \leq 0$ **then**
    roll back to s_t; stall $\leftarrow$ stall $+1$; **if** stall $> \varrho$ **then break**
  else
    $s_{t+1} \leftarrow$ state of $\arg\max_{d \in \mathcal{K}} U'_d$; stall $\leftarrow 0$; push s_{t+1} to the state pool
  end if
end for
return best verified state visited, with its decoded edit trace

Algorithm 2 One EPIFORGE training cycle

Require: replay buffer $\mathcal{B}$ of verified transitions, state pool $\mathcal{S}$, coefficients $\kappa, \kappa_{\text{pen}}$

offline: fit ECHO by minimizing $\mathcal{L}_{\text{Echo}}$ of eq. (5) on $\mathcal{B}$
policy: for each state s sampled from $\mathcal{S}$
  draw G genotypes against the same (A, E); decode by eq. (4)
  $r_i \leftarrow \hat{U}(s, a_i) - \kappa_{\text{pen}}\hat{\sigma}(s, a_i)$, replaced by the realized U where the child was verified
  $\hat{A}_i \leftarrow (r_i - \text{mean}_j r_j)/(\text{std}_j r_j + \varepsilon)$
  one clipped GRPO update on $\mathcal{L}_\pi$ with KL to the frozen reference π_{ref}
online: run algorithm 1 from states in $\mathcal{S}$; select top-k by the upper confidence bound; verify with $\mathcal{V}$; accept or roll back
bookkeeping: push all verified transitions, accepted or rolled back, to $\mathcal{B}$; push accepted states to $\mathcal{S}$
refresh: update the standardization statistics of eq. (2); re-fit ECHO on the enlarged $\mathcal{B}$; anneal κ
return next cycle

residue in the reference complex, retained only when it has at least five residues. Chains with unresolved CDR residues, non-standard residues in a CDR, or a missing canonical disulfide are dropped; residues missing a carbonyl oxygen are retained and handled by the fallback of eq. (7). Redundancy control operates on connected components of the union of the three relations described in the main text, and components are assigned whole. Clustering on the antibody alone would be insufficient: a test target whose antigen appears in training with a different antibody is not a fair generalization test, because the epitope and its complementary chemistry are then already known. Nanobody entries are identified by the absence of a light chain and kept as a separate subset rather than pooled, since their loop geometry and interface size differ systematically. The offline pairs used to pre-fit ECHO are generated by applying decoded mutation sets sampled uniformly at random from the legal space of eq. (4) to training-split parents and verifying both parent and child, so that the surrogate sees the same action distribution the untrained policy will produce.

A.11 Ablation protocol

Ablations reuse a single checkpoint and perform inference only. Each target is given one initializer root shared across configurations, and predictions are cached and keyed by sequence, so a design reached by two configurations is predicted once. Each configuration exports one design per target and only exported designs are evaluated under the reporting panel, so reporting-panel calls equal configurations times targets and do not grow with search width or depth. The module ablation rows of table 3 are not budget-matched to the depth rows, which we state in the main text rather than in a footnote because the two differ in how many candidates are verified as well as in what is disabled.

A.12 Implementation and hyperparameters

The four quantities EpiForge introduces beyond a standard policy-optimization setup are the optimism coefficient κ of section 3.6, the pessimism coefficient κ_{pen} of eq. (6), the verification width k, and the step budget T. All components are open and installable through a package manager or container. Structure prediction uses openly released inference code and weights, and numbering, surface computation, contact and clash detection, DockQ, descriptor computation, liability matching, and humanness scoring use openly available libraries. No licensed molecular-modeling suite is used at any point and no physics-based repacking or minimization is performed; where such a step would conventionally appear we substitute a fast learned antibody structure predictor. Values for each hyperparameter group, together with the canonical amino-acid alphabet order used by eq. (4), versions and container digests, are released with the code.

A.13 Affinity maturation

Table 6: **Affinity maturation** under a bounded mutation budget. Improved is the fraction of proposals predicted to improve binding. The remaining columns are predicted energetic terms, lower being better. These are predicted rather than measured quantities and the methods differ in how many mutations each is permitted, so we regard the comparison as unresolved.

Method	Improved (%) ↑	ΔInteraction ↓	ΔvdW clash ↓	ΔElectrostatics ↓	ΔSC H-bond ↓	Interface clash ↓
RefineGNN	0.000	30.919	23.229	2.977	3.052	8.857
dyMEAN	0.000	32.830	25.517	3.651	**2.497**	9.493
AbODE	0.000	59.666	62.498	−0.274	−0.271	21.200
AbDPO	0.000	59.666	62.498	−0.274	−0.271	21.200
EpiForge (binding head)	**0.292**	**29.979**	**21.667**	**1.582**	2.586	**7.907**

Three observations. EpiForge is the only method that returns an improved design at all, and it leads every energetic column except side-chain hydrogen bonding, where dyMEAN is lower; the negative electrostatic and hydrogen-bond entries of the last two rows are not marked, for the reason given below. All four baselines improve no design under this protocol, which says as much about the protocol's bounded mutation budget as about those methods, since a co-design model asked to change a handful of residues is operating far from the regime it was trained for — and by the same token our own margin here is partly a statement about fit to the protocol. The AbODE and AbDPO rows are identical across every column, which we flag as a pipeline issue to be resolved rather than presenting as a finding; until it is resolved, neither row should be read as a measurement of that method.

A.14 Additional results

Table 7 gives the CDR-H3 length and held-out breakdown discussed in section 4.2. This section further contains the per-family breakdown underlying it; the realized acceptance and rollback rates per panel, with the fraction of episodes terminated by the stall rule; the calibration curve of the Echo variance head against realized error, including the effect of the independence assumption used to form the scalar $\hat{\sigma}$; the distribution of decoded operator labels over accepted and rolled-back edits, which is the audit trail the labels of Appendix A.6 exist to provide; the distribution of decoded positions over the CDRs, which under eq. (4) is determined by the policy rather than by any prior and which we therefore measure rather than assume; the agreement between the in-loop verifier and each reporting predictor on the same designs, which bounds how much of the in-loop gain is expected to transfer; and per-panel results on the binding datasets reported separately rather than pooled, since pooling

Table 7: **Breakdown by CDR-H3 length and on held-out panels.** Base is dyMEAN, the strongest baseline on overall CDR-H3 recovery; RMSD is the mean over the two reporting predictors. Subsets differ greatly in size, so the count column should be read alongside them.

	CDR-H3 recovery (%)↑			CDR-H3 RMSD (Å)↓			
Category	Base	EPIFORGE	Δ	Base	EPIFORGE	Δ	Count
CDR-H3 short (≤ 9)	58.73	53.77	−4.96	4.644	0.986	−3.658	23
CDR-H3 medium (10 to 15)	47.55	49.00	+1.45	6.554	2.890	−3.664	209
CDR-H3 long (> 15)	29.74	37.16	+7.42	10.626	6.248	−4.378	112
RAbD (held out)	40.73	49.84	+9.11	3.959	3.049	−0.910	60
CoV-AbDab (held out)	33.41	42.40	+8.99	5.101	4.335	−0.766	60
Overall (test)	40.08	**45.32**	+5.24	6.821	4.104	−2.717	344

across assays with different dynamic ranges would produce a correlation reflecting assay identity as much as design quality.