19 DEGREES OF FREEDOM BEAT FULL-PARAMETER FINE-TUNING IN FIXED-BACKBONE ANTIBODY DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Inverse-folding models generate antibody CDR sequences compatible with antibody–antigen structures, but are trained for structural compatibility rather than binding. We introduce PACT (Position-Additive Composition Tilt), which steers a frozen decoder with a shared amino-acid bias of only 19 effective degrees of freedom. Its coefficients are estimated in closed form from within-backbone reward differences, requiring no gradient training. Across three antigens, PACT improves mean Rosetta interface reward by 0.32–0.63 and increases shape complementarity on every target. In contrast, advantage regression yields no significant reward gain over the frozen reference, while filtered SFT reduces reward and falls behind PACT by 0.99–1.83 units. To investigate the limited benefit of full-parameter training, we examine how it reweights candidate designs. Across six tested objectives, fine-tuning produces log-likelihood ratios relative to the frozen reference that are nearly uncorrelated with reward differences among candidates for the same backbone. Checkpoint analysis of advantage regression further shows that these candidates retain highly aligned gradients throughout training. The fitted PACT biases also transfer across targets and improve ESM-IF and AntiFold without refitting. Together, these findings support direct steering of residue preferences as an effective and transferable alternative to the tested full-parameter updates.

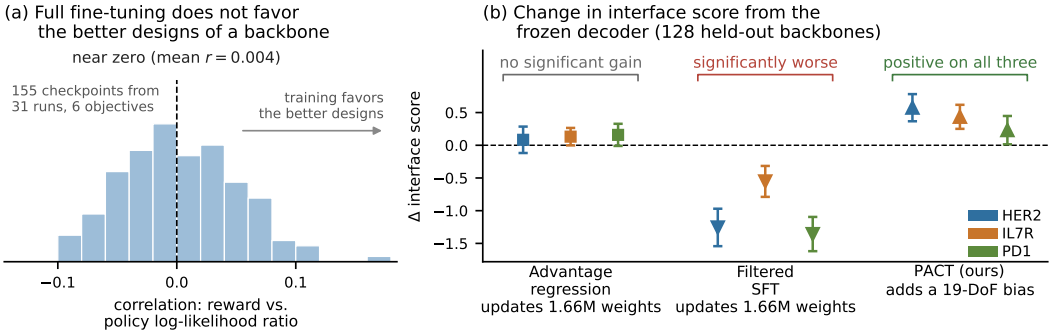


Figure 1: **Full-parameter fine-tuning fails to favor higher-reward designs, while a 19-DoF tilt improves held-out reward.** (a) Within-backbone reward and policy log-likelihood ratio are nearly uncorrelated in the six-objective training sweep. (b) Change in refolded interface score on 128 held-out backbones, with up to eight designs each, versus the frozen decoder (95% CIs).

1 INTRODUCTION

Inverse-folding models can generate many complementarity-determining region (CDR) sequences compatible with a fixed antibody backbone (Dreyer et al., 2023; Høie et al., 2025), but only a limited number can be experimentally tested (Hie et al., 2024). Since these models are trained for structural compatibility rather than antigen binding, their highest-probability designs need not be the most promising binders. The challenge is therefore to use binding-related feedback to shift generation toward better interfaces while preserving the structural prior learned by the pretrained decoder.

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.

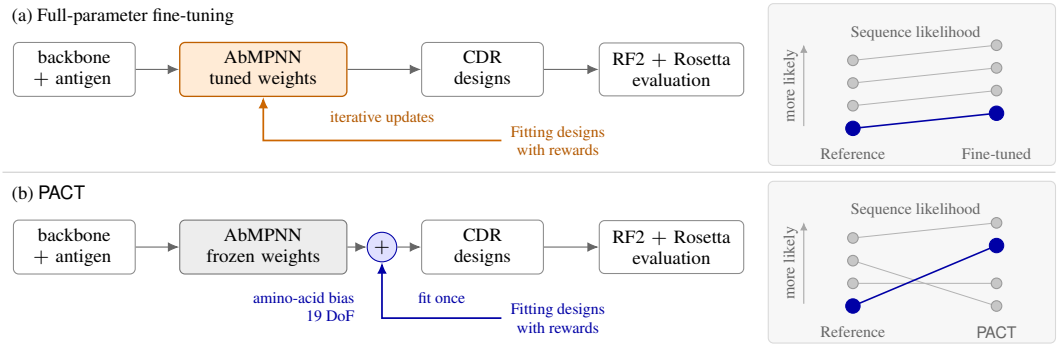


Figure 2: **Full-parameter fine-tuning versus PACT.** Both use fitting designs with precomputed rewards: **(a)** iteratively updates model weights; **(b)** fits a shared amino-acid logit bias once, keeping the decoder frozen. Horizontal paths show held-out generation and evaluation; evaluation scores do not update weights or bias. Right: schematic likelihoods; blue marks the highest-reward design.

A natural approach is to use reward supervision to fine-tune the full decoder (Xu et al., 2023). Here, the reward is a Rosetta interface score calibrated to experimental $\Delta\Delta G$ measurements from SKEMPI v2.0 (Jankauskaitė et al., 2019). Under the standard KL-regularized objective, higher-reward designs for the same backbone should gain more likelihood relative to the reference policy (Rafailov et al., 2023). In our experiments, however, full-parameter fine-tuning does not reliably produce this relative reordering. Instead of selectively increasing the likelihood of higher-reward candidates within a backbone, training mainly shifts candidate likelihoods together. Across six objectives and learning rates spanning two orders of magnitude, the resulting relative likelihood changes remain nearly uncorrelated with within-backbone reward differences.

To characterize this behavior in a representative full-parameter objective, we examine advantage regression in more detail. Reward-dependent reweighting requires higher- and lower-reward candidates from the same backbone to receive different likelihood updates. However, their gradients remain highly aligned throughout full-parameter fine-tuning, while parameter updates tend to move their likelihoods in similar directions rather than separate them according to reward. Our local update analysis captures these likelihood changes, but the relative shifts remain poorly aligned with reward. We also evaluate held-out designs after independent refolding. Advantage regression yields no significant reward improvement on any target. Filtered SFT shows a complementary failure: although it closely fits the selected high-reward training designs, it lowers reward on held-out backbones.

These results suggest that the tested full-parameter objectives provide an ineffective way to express the reward preferences that distinguish candidates for the same backbone. We therefore take a more direct approach: estimate these preferences from reward differences and inject them at the decoder output rather than learning them through updates to millions of parameters. We introduce PACT (Position-Additive Composition Tilt), which freezes the decoder and adds a shared amino-acid bias to its output logits. The bias has only 19 effective degrees of freedom and is estimated in closed form from within-backbone reward differences, requiring no gradient training (Figure 2).

Across three antigens with experimentally determined epitopes (HER2, IL7R, and PD1), PACT improves interface reward on most held-out backbones (Figure 4) and increases shape complementarity on all three targets. Shuffling the fitted coefficients removes these gains, while pooled biases transfer across targets and improve ESM-IF and AntiFold without refitting. To test whether these gains extend beyond the fitting reward and fixed structural pose, we separate fitting from evaluation. IgFold (Ruffolo et al., 2023) provides a reproducible fitting signal while preserving the designed antibody–antigen pose, whereas held-out complexes are independently re-predicted with antibody-adapted RoseTTAFold2 (Bennett et al., 2025). We also report shape complementarity separately to assess interface packing beyond the aggregate reward.

Contributions. First, we introduce PACT, a 19-DoF method that steers a frozen antibody decoder through a shared amino-acid bias estimated in closed form from within-backbone reward differences. Second, across six tested full-parameter objectives, we find little reward-aligned reweighting and use gradient analysis of advantage regression to characterize how likelihood updates differ among same-backbone candidates. Third, PACT improves held-out reward and shape complementarity

across three antigens, achieves higher mean reward than the tested full-parameter baselines, and transfers across targets and to ESM-IF and AntiFold without refitting.

2 PRELIMINARIES

Let the context $C := (B, S_{\mathcal{I}_{\mathrm{FR}}}, \mathcal{I}_{\mathrm{CDR}})$ denote a fixed antigen–antibody backbone B, its framework sequence, and the set of designable CDR positions (Murphy & Weaver, 2016). A policy $\pi_\theta(\cdot \mid C)$ generates a sequence $\hat{S}$ over the $|\mathcal{A}| = 20$ canonical amino acids while preserving the framework. Our reference policy π_{ref} is ProteinMPNN (Dauparas et al., 2022) with pretrained AbMPNN weights (Dreyer et al., 2023).

2.1 KL-REGULARIZED POLICY OPTIMIZATION

To favor higher-reward designs while retaining the reference distribution, we consider the standard KL-regularized objective

$$\max_{\pi_\theta} \mathbb{E}_{C\sim D,\ \hat{S}\sim\pi_\theta(\cdot|C)} \left[r(C,\hat{S}) - \beta\, \mathrm{KL}(\pi_\theta(\cdot \mid C) \,\|\, \pi_{\mathrm{ref}}(\cdot \mid C)) \right], \tag{1}$$

where $\beta > 0$ controls regularization toward the pretrained reference distribution. Over unrestricted conditional distributions, the optimum has the Boltzmann form (Rafailov et al., 2023)

$$\pi^*(\hat{S} \mid C) \propto \pi_{\mathrm{ref}}(\hat{S} \mid C) e^{r(C,\hat{S})/\beta}, \qquad V^*(C) = \beta \log \mathbb{E}_{\hat{S}\sim\pi_{\mathrm{ref}}(\cdot|C)} \left[e^{r(C,\hat{S})/\beta} \right]. \tag{2}$$

Appendix A gives the derivation. The optimum specifies how rewards should reweight the reference distribution. We next examine whether full-parameter fine-tuning achieves this reweighting.

2.2 SURROGATE REWARD MODEL

We score designs using three Rosetta interface metrics: shape complementarity (SC) measures geometric fit, buried surface area (ΔSASA) measures interface size, and interface hydrogen bonds (Hbonds) capture polar contacts. We fit their weights by regressing experimental $\Delta\Delta G$ on mutant–wild-type changes in these metrics in the antibody–antigen subset of SKEMPI v2.0 (Jankauskaitė et al., 2019), then reverse the signs to reward favorable changes:

$$R = 12.97\,\mathrm{SC} + 38.26\,\frac{\Delta\mathrm{SASA}}{9200} + 11.05\,\frac{\mathrm{Hbonds}}{394}. \tag{3}$$

The weights and normalization constants are fixed across targets. We write $r(C,\hat{S})$ for this score evaluated on the predicted complex for design $\hat{S}$ in context C. Sections 4 and 5.1 specify the structure-prediction pipelines used for fitting and held-out evaluation. Regression details and coefficient uncertainties are in Appendix D.3, and the unrounded score is in Appendix D.6.

3 DIAGNOSING FULL-PARAMETER FINE-TUNING

We now ask whether full-parameter fine-tuning actually reweights same-backbone candidates according to reward. The KL-regularized optimum gives a simple target: higher-reward candidates should receive larger likelihood ratios relative to the reference. We use this relationship to define a diagnostic objective and then examine how closely training follows it.

How reward should change likelihoods. Under the KL-regularized optimum in Equation (2), reward differences within a backbone translate directly into likelihood-ratio differences. Define the optimal advantage by

$$A(C,\hat{S}) := r(C,\hat{S}) - V^*(C) = \beta \log \frac{\pi^*(\hat{S} \mid C)}{\pi_{\mathrm{ref}}(\hat{S} \mid C)}.$$

This suggests fitting the policy so that its realized log-likelihood ratio matches the observed advantage:

$$\mathcal{L}(\pi_\theta) = \mathbb{E}_{C\sim D}\mathbb{E}_{\hat{S}\sim\pi_{\mathrm{ref}}(\cdot|C)} \left[\left(\beta \log \frac{\pi_\theta(\hat{S} \mid C)}{\pi_{\mathrm{ref}}(\hat{S} \mid C)} - \left(r(C,\hat{S}) - \hat{V}(C) \right) \right)^2 \right]. \tag{4}$$

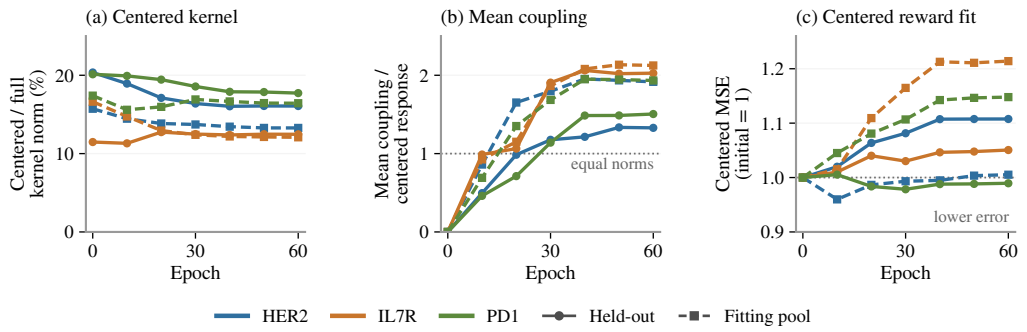


Figure 3: **Shared gradients persist, but reward fitting barely improves.** (a) Fraction of gradient-kernel norm remaining after centering. (b) Mean-shift effect relative to the reward-dependent response in Equation (5). (c) Centered reward-fit MSE normalized to initialization. Curves summarize 12 backbones per antigen/split with 16 fixed candidates each. Details are in Appendices D.9 and E.4.

The baseline $\hat{V}(C)$ is shared across candidates from the same backbone. If $\hat{V}(C) = V^*(C)$, the Boltzmann policy π^* achieves zero loss in Equation (4) and is therefore a global minimizer over unrestricted policies. In practice, we consider two baselines: a sample log-mean-exp reward baseline motivated by Equation (2), and the mean reward within each backbone. The latter centers sampled rewards but does not in general estimate $V^*(C)$. Experiments apply the same squared residual to fixed offline candidates. Generation and scoring conventions are specified in Appendices D.4 and D.9.

We next ask whether full-parameter training realizes the within-backbone reweighting predicted above. Across six tested objectives, changes in policy log-likelihood ratios remain nearly uncorrelated with reward differences within the same backbone, with a mean correlation of 0.004 across 31 runs and 155 recorded epochs (Figure 7). This behavior is observed across the tested learning rates, gradient-clipping thresholds, β values, and baseline choices. Objective definitions and diagnostic results are reported in Appendices D.8 and E.1.

How fine-tuning changes relative likelihoods. To understand why fine-tuning produces little reward-aligned reweighting, we examine how each parameter update changes the relative likelihoods of candidates from the same backbone. Since higher-reward candidates must move relative to the others rather than simply shift with them, we analyze the centered log-ratio update induced by the advantage-matching objective. For a fixed context and decoding order with n candidates, let

$$g_i = \nabla_\theta \log \pi_\theta(\hat{S}_i \mid C), \qquad K_{ij} = \langle g_i, g_j \rangle, \qquad \ell_i = \beta \log \frac{\pi_\theta(\hat{S}_i \mid C)}{\pi_{\text{ref}}(\hat{S}_i \mid C)}.$$

Let $P = I - \mathbf{1}\mathbf{1}^\top/n$, so $P\ell$ isolates within-backbone differences by removing the mean log-ratio shift. Under the mean-reward baseline, $A_i = r(C, \hat{S}_i) - \hat{V}(C)$ is also centered, with $PA = A$. Writing $e_t = A - \ell_t$ and denoting the mean log-ratio by $\bar{\ell}_t$, an SGD step of size h_t gives

$$P\Delta\ell_t = \underbrace{\eta_t P K_t P e_t}_{\text{centered-residual response}} \underbrace{- \eta_t \bar{\ell}_t P K_t \mathbf{1}}_{\text{mean coupling}} + P R_t, \qquad \eta_t = 2h_t\beta^2/n. \tag{5}$$

R_t is the Taylor remainder. The first term captures how reward-dependent residuals drive candidates apart through the centered gradient kernel. The second couples relative updates to the backbone-level mean likelihood shift. See Appendix B for the derivation and extension to optimizer updates.

Figure 3 tracks the quantities above throughout advantage-regression training. Same-backbone gradients remain highly aligned through epoch 60, when centering retains only 12–18% of the kernel norm across antigen/split medians. Meanwhile, the backbone-wide mean shift becomes increasingly important relative to the reward-dependent response. Reward fitting barely improves: five of six groups end no better than initialization, and the remaining group improves only marginally.

The first-order decomposition also closely matches observed likelihood changes under controlled SGD and AdamW updates (Appendix E.5). A separate held-out refolding comparison tests design

quality: advantage regression yields no significant reward gain, while filtered SFT lowers reward relative to the frozen AbMPNN reference (Table 1 and Appendix E.1). Together, these results motivate keeping the decoder fixed and expressing reward preferences directly through a low-dimensional amino-acid bias (Section 4).

4 PACT: POSITION-ADDITIVE COMPOSITION TILT

The analysis above suggests that full-parameter fine-tuning is a poor mechanism for expressing within-backbone reward preferences: candidate likelihoods tend to move together rather than separate according to reward. We therefore ask whether these preferences can be injected directly at the decoder output instead. PACT does this by freezing the decoder and adding a shared amino-acid bias $b = (b_a)_{a\in\mathcal{A}}$ at every designable CDR position.

From reward to a composition tilt. To motivate the shared bias, consider an additive approximation in which each amino-acid identity contributes a position-independent reward term. Specifically,

$$r(C, \hat{S}) \approx r_0(C) + \sum_{q\in\mathcal{I}_{\mathrm{CDR}}} b_{\hat{S}_q}. \tag{6}$$

Substituting it into the Boltzmann optimum in Equation (2) gives

$$\pi_b(\hat{S} \mid C) \;\propto\; \pi_{\mathrm{ref}}(\hat{S} \mid C)\, \exp\Big(\tfrac{1}{\beta} \sum_{q\in\mathcal{I}_{\mathrm{CDR}}} b_{\hat{S}_q}\Big), \qquad b \in \mathbb{R}^{|\mathcal{A}|}. \tag{7}$$

The backbone-dependent term $r_0(C)$ is absorbed into the normalizer, leaving a sequence-level tilt determined only by amino-acid composition. This directly motivates a shared residue bias on the frozen decoder.

To implement this composition preference in the autoregressive decoder, PACT adds b_a/β to the frozen decoder logit z_a for amino acid a at each designable position q. The resulting local sampling distribution is

$$\pi_b^{\mathrm{loc}}(\cdot \mid \hat{S}_{<q}, C) = \mathrm{softmax}\left(\frac{z(\hat{S}_{<q}, C) + b/\beta}{\tau}\right). \tag{8}$$

Here, $1/\beta$ controls the strength of the bias and τ is the decoding temperature. A common shift of all coefficients cancels under the softmax, leaving 19 effective degrees of freedom (Appendix C.2). Because the normalization is applied separately at each decoding step and can depend on the prefix, this local rule is not generally identical to the sequence-level tilt in Equation (7). We give the relationship in Appendix C.1. Decoding settings are provided in Appendix D.4.

To isolate within-backbone reward differences, we center rewards by subtracting the mean fitting reward $\bar{r}(C)$ for each backbone and restrict attention to CDR positions $\mathcal{V}(C)$ that vary across sampled designs. For each amino acid a, we then estimate its bias by averaging the centered reward over designs in which a appears at a variable position:

$$\hat{b}_a \;=\; \frac{1}{|\mathcal{N}_a|} \sum_{(C,\hat{S},q)\in\mathcal{N}_a} \big(r(C, \hat{S}) - \bar{r}(C)\big), \qquad \mathcal{N}_a = \big\{(C, \hat{S}, q) : q \in \mathcal{V}(C),\ \hat{S}_q = a\big\}. \tag{9}$$

Thus, $\hat{b}_a$ captures the empirical association between amino acid a and reward within the same backbone. Because each occurrence is attributed the reward of its full sequence, the estimator does not disentangle contributions from co-occurring residues and should not be interpreted as an exact additive reward-model fit (Appendix C.3). Algorithm 1 summarizes the complete procedure.

Reward estimates for fitting. We estimate the bias coefficients from rewards of sequences sampled from the frozen reference. For each fitting backbone, we predict each sampled antibody with IgFold (Ruffolo et al., 2023). We use IgFold at this stage because it provides a reproducible antibody structure prediction while allowing us to preserve the designed antibody–antigen pose during reward fitting. We align the predicted antibody framework to the designed complex while keeping the antigen fixed, then compute the reward in Equation (3) using Rosetta's `InterfaceAnalyzer`. This measures sequence-dependent changes under a common designed pose. Held-out evaluation instead re-predicts the full complex independently (Section 5.1).

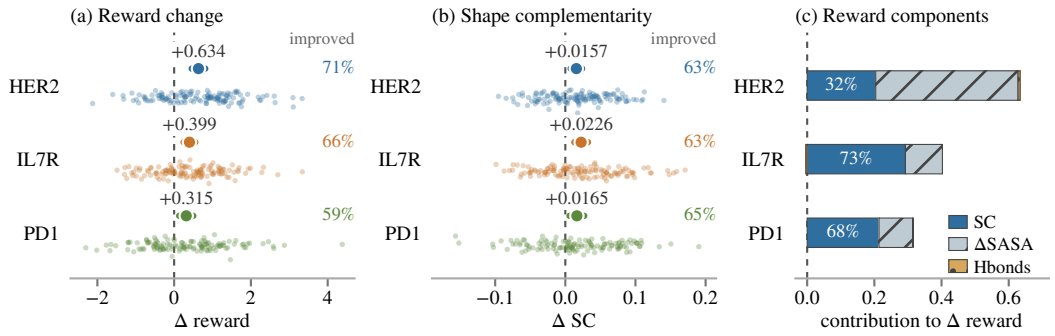


Figure 4: **PACT improves reward and interface packing on held-out backbones.** (a,b) Per-backbone mean PACT–AbMPNN differences, with overall mean, 95% CI, and fraction improved. (c) Decomposition of the reward gain where labels show SC's contribution. Full results are in Table 12.

5 Experiments

We evaluate PACT after full-complex refolding, compare it with full-parameter fine-tuning, and test whether its gains extend beyond the reward used for fitting. We then use control and transfer experiments to ask whether the learned amino-acid preferences are meaningful and whether they generalize to other inverse-folding models without refitting.

5.1 Experimental Setup

Targets and fitting data. We study **HER2**, **IL7R**, and **PD1**, using epitopes from solved complexes (PDB 1S78, 6P50, and 5GGS). For each target, RFantibody (Bennett et al., 2025) generates 256 backbones, split into 128 for fitting and 128 for evaluation. We fit PACT on 16 AbMPNN designs per fitting backbone, scored with the IgFold pipeline proposed in Section 4. Each amino-acid occurrence at a variable CDR position is treated as one observation, so multiple occurrences within the same design are counted separately. We estimate a coefficient only for amino acids observed at least 150 times; coefficients for less frequent residues are set to zero. This leaves 17–18 nonzero coefficients per target (Appendix F.2). Evaluation backbones are excluded from coefficient fitting.

Decoding and comparison budgets. AbMPNN is decoded nearly greedily at $\tau = 10^{-4}$, with diversity introduced by random decoding order. The main PACT–AbMPNN comparison uses 16 designs per held-out backbone, while method comparisons use matched sets of eight designs per backbone. Each comparison uses unsteered AbMPNN outputs generated under the same decoding budget as its reference. Bias-strength settings and control-specific scaling are reported in Appendix D.4.

Refolding and evaluation. To test whether improvements persist beyond the fixed pose used for coefficient fitting, we re-predict each held-out antibody–antigen complex with antibody-adapted RoseTTAFold2 (RF2). This allows both the relative antibody–antigen pose and the local interface geometry to change. We then recompute the reward in Equation (3) and evaluate shape complementarity (SC) separately to distinguish improved interface packing from simply forming a larger interface. All methods are evaluated with the same RF2 and Rosetta pipeline (Appendices D.6 and F.9).

Statistical comparisons. All method comparisons are paired at the backbone level: methods use the same backbones, designable positions, and planned decoding budgets, and available designs are first averaged within each backbone before computing paired differences. We report 95% confidence intervals and perform two-sided tests across backbones. Primary reward and SC comparisons are corrected across the three antigens, as are reward comparisons against each full-parameter baseline. Decoder-transfer tests are corrected across six comparisons. Other metrics and steering controls are treated as exploratory. Test details, including incomplete outputs, are in Appendix D.5.

5.2 Held-Out Reward and Interface Packing

After independent RF2 refolding, PACT improves the interface reward on all three targets (Figure 4a), with mean gains of 0.32–0.63 and improvements on 59–71% of held-out backbones. Shape complementarity also increases on all three targets (Figure 4b), showing that the reward gains are accompanied by better geometric packing rather than interface expansion alone. Both reward and SC improvements remain significant after correction across the three targets. The reward decomposition in Figure 4c shows what drives these gains. SC accounts for 32–73% of the total reward improvement. HER2 gains are distributed across all three reward components, whereas IL7R is driven primarily by improved SC despite fewer interface hydrogen bonds. For PD1, SC is the only component with a statistically resolved increase (Table 12).

Table 1: **Comparison of reward-steering methods on held-out backbones.** Mean changes relative to frozen AbMPNN across 128 backbones, with up to eight designs per backbone. Bold denotes the best mean among AbMPNN-based methods; arrows indicate favorable directions. p_R gives unadjusted two-sided backbone-level reward tests versus AbMPNN. E is unrelaxed Rosetta separated energy (REU); scRMSD is in Å. Statistical corrections and coverage are reported in Appendix D.5.

		AbMPNN-based methods			Sequence-only references	
Target	Metric	**PACT**	Advantage regression	Filtered SFT	AbLang (Olsen et al., 2022)	nanoBERT (Hadsund et al., 2024)
HER2	$\Delta R \uparrow$	$\mathbf{+0.574}$	$+0.084$	-1.256	-0.318	$+0.058$
	p_R	5.7×10^{-8}	0.42	7.7×10^{-18}	0.068	0.75
	ΔSC $\uparrow$	$\mathbf{+0.0149}$	$+0.0010$	-0.0527	-0.0096	$+0.0084$
	$\Delta E \downarrow$	$\mathbf{-424}$	-100	$+305$	$+379$	-113
	ΔscRMSD $\downarrow$	$\mathbf{-0.022}$	-0.008	$+0.372$	$+0.803$	$+0.510$
IL7R	$\Delta R \uparrow$	$\mathbf{+0.435}$	$+0.132$	-0.552	-0.180	-0.027
	p_R	3.5×10^{-6}	0.056	4.7×10^{-6}	0.16	0.84
	ΔSC $\uparrow$	$\mathbf{+0.0265}$	$+0.0101$	-0.0392	-0.0101	-0.0040
	$\Delta E \downarrow$	$\mathbf{-430}$	-122	$+2192$	$+556$	-14
	ΔscRMSD $\downarrow$	$\mathbf{-0.041}$	-0.023	$+0.548$	$+0.993$	$+0.690$
PD1	$\Delta R \uparrow$	$\mathbf{+0.231}$	$+0.159$	-1.357	-0.360	-0.483
	p_R	0.037	0.068	2.9×10^{-24}	0.035	2.1×10^{-3}
	ΔSC $\uparrow$	$\mathbf{+0.0125}$	$+0.0085$	-0.0730	-0.0050	-0.0237
	$\Delta E \downarrow$	$\mathbf{-279}$	-100	$+1386$	$+25$	$+653$
	ΔscRMSD $\downarrow$	$\mathbf{-0.046}$	-0.030	$+0.512$	$+0.855$	$+0.661$

5.3 Comparison with Full-Parameter Fine-Tuning

Having established that PACT improves over the frozen decoder, we next compare it with two full-parameter alternatives that update all 1.66M AbMPNN weights: advantage regression (Equation (4)) and filtered SFT on the highest-reward quartile within each fitting backbone. Using the same eight-design evaluation budget, PACT achieves the highest mean reward and SC of the three adaptation methods on every target. The same trend holds for separated energy and heavy-chain self-consistency RMSD (scRMSD), two evaluation metrics not included in the fitting reward (Table 1).

The clearest difference is in interface reward. PACT exceeds filtered SFT by 0.99–1.83 reward units and improves 73–87% of held-out backbones, with significant gains on all three targets after correction. PACT has a higher mean reward than advantage regression on all three targets, with pairwise differences significant after correction on HER2 and IL7R. More broadly, the full-parameter baselines show no significant reward improvement over the frozen reference: advantage regression shows no significant reward gain on any target, while filtered SFT reduces both reward and SC despite closely fitting its high-reward training designs (Appendix E.1).

On PD1, the eight-design comparison gives a positive PACT–AbMPNN reward gain ($+0.231$) but does not pass the three-target correction ($p = 0.037$). The primary 16-design evaluation on the same backbones yields a more precise estimate and detects a significant gain.

5.4 Structural Metrics Beyond the Fitted Reward

Here we focus on two additional structural metrics that are not included in the reward: Rosetta separated energy and heavy-chain self-consistency RMSD (scRMSD). PACT lowers separated energy on all three targets relative to AbMPNN, with exploratory 95% CIs below zero in the 16-design comparison (Figure 5a). This provides an out-of-objective indication of improved Rosetta energetics, although separated energy remains an unrelaxed structural score rather than an estimate of binding free energy. Heavy-chain scRMSD measures how closely the refolded antibody backbone matches the designed backbone after heavy-chain alignment. PACT shows no statistically resolved scRMSD change from AbMPNN, while reducing scRMSD relative to filtered SFT by 0.39–0.59Å across all three targets (Figure 5b). Its reward gains therefore come without a detected loss of backbone fidelity.

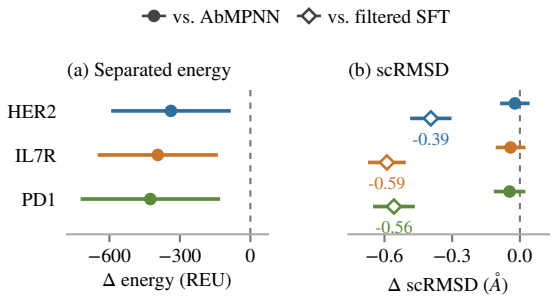


Figure 5: PACT lowers separated energy with no detected loss of backbone consistency. (a) Separated-energy differences vs. AbMPNN at the 16-design budget. (b) scRMSD differences vs. AbMPNN (circles) and filtered SFT (diamonds) at the eight-design budget. Lower is better; bars show exploratory 95% CIs. See Appendix F.1.1.

5.5 Testing the Learned Residue Preferences

To test whether PACT's gains depend on the learned amino-acid preferences rather than simply the magnitude of the bias, we compare the fitted vector with two controls: a shuffled version of the same coefficients and a side-chain-volume prior (Table 2). With bias magnitude and steering strength matched, the learned residue assignments make the difference: the fitted PACT vector improves reward on all three targets, whereas shuffling the same coefficient values produces negative mean changes throughout. Thus, bias magnitude alone is insufficient to recover the gain. A side-chain-volume prior also improves reward on HER2 and IL7R, but provides essentially no gain on PD1, where PACT exceeds it by 0.286 reward units ($p = 0.006$). Together, these controls show that the learned residue assignments matter, with a resolved benefit beyond residue size on PD1. Full control results are in Tables 19 and 20.

Table 2: **Learned residue preferences support both steering and reuse.** Mean reward changes from AbMPNN across 128 backbones with eight designs each. Shading marks PACT; bold indicates the larger point estimate. Controls are L_1-matched, while pooled coefficients retain the source norm.

Bias vector	HER2 ΔR	IL7R ΔR	PD1 ΔR
PACT: target-specific	**+0.786**	+0.306	+0.287
PACT: pooled from other targets	+0.410	**+0.639**	**+0.328**
Shuffled coefficients	−0.760	−0.097	−0.174
Side-chain volume	+0.754	+0.424	+0.001

5.6 Reusing Learned Residue Preferences

The controls above show that the learned residue assignments matter. We next ask whether these preferences are specific to the target and decoder used for fitting, or can be reused without refitting.

Reusing coefficients across targets. We next test whether the learned residue preferences transfer across antigens without target-specific fitting. For each target, we pool the bias vectors fitted on the other two antigens and apply the resulting bias directly to AbMPNN. Despite using no data from the target antigen, the transferred bias improves reward on all three targets and yields significant SC gains on two (Tables 2 and 19). Its reward is comparable to the target-specific fit overall, exceeding it on IL7R, trailing it on HER2, and showing no resolved difference on PD1 (Table 20). These results indicate that a substantial part of the learned residue preference is shared across targets and can be reused without refitting.

Transferring to other decoders. We next apply the same AbMPNN-derived pooled biases to frozen ESM-IF (Hsu et al., 2022) and AntiFold (Høie et al., 2025), without refitting either the bias or decoder. Relative to each decoder's unsteered outputs, reward improves by 0.40–0.85 units in all six target–decoder combinations, with all gains significant after correction (Figure 6). The same AbMPNN-derived biases improve reward across both ESM-IF and AntiFold without refitting, indicating that the learned residue preferences are not specific to a single decoder or target. The transfer gains arise mainly from increased buried interface area, while shape-complementarity changes remain unresolved.

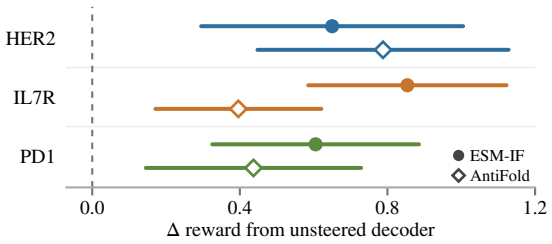


Figure 6: PACT transfers across decoders without refitting. Pooled AbMPNN-derived biases improve reward in all six target–decoder combinations. Points show mean gains over unsteered baselines with 95% CIs (Table 23).

6 Related Work

Fixed-backbone antibody design. Inverse-folding models generate sequences compatible with a fixed backbone, including ProteinMPNN (Dauparas et al., 2022), AbMPNN (Dreyer et al., 2023), AntiFold (Høie et al., 2025), and ESM-IF (Hsu et al., 2022). These models primarily learn sequence–structure compatibility rather than antigen binding. Joint sequence–structure methods instead optimize the backbone as well (Jin et al., 2022; Kong et al., 2023a; Luo et al., 2022; Kong et al., 2023b; Zhou et al., 2024; Chai Discovery Team et al., 2025). Antibody LMs support sequence modeling and design (Olsen et al., 2022; Leem et al., 2022; Shuai et al., 2023; Hadsund et al., 2024; Hie et al., 2024; Lee et al., 2025), but sequence-only models do not directly condition on antigen structure.

Reward-based steering. Reward optimization has been applied broadly to biological sequence generation (Angermueller et al., 2020; Chen et al., 2023; Wang et al., 2025), including protein design (Xu et al., 2023; Xue et al., 2025; Park et al., 2024; Wang et al., 2026). These methods typically adapt model parameters using reward supervision. In the full-parameter objectives we test, however, training produces little reward-aligned reweighting among candidates for the same backbone. PACT instead keeps the decoder frozen and estimates a shared residue bias in closed form.

Structure-based evaluation. Structure predictors such as IgFold (Ruffolo et al., 2023), antibody-adapted RoseTTAFold2 (Bennett et al., 2025), and Boltz-2 (Passaro et al., 2025) provide complementary ways to evaluate designed antibodies. We separate the predictor used for fitting from independent full-complex refolding at evaluation. Protocol details are given in Appendix F.9.

7 Conclusion

Our results suggest that effective reward steering need not require a more expressive update. Across six tested full-parameter objectives, fine-tuning produces little reward-aligned reweighting among designs for the same backbone. Our training-time analysis further shows that same-backbone gradients remain highly aligned and that relative likelihoods continue to fit reward differences poorly. These observations point to a mismatch between the capacity of the decoder and the mechanism used to express fine-grained reward preferences.

PACT takes a different approach: it leaves the decoder fixed and makes reward-associated residue preferences directly available at generation time. With only 19 effective degrees of freedom and no gradient training, PACT improves held-out interface reward and packing across three antigens, outperforms filtered SFT on all three targets and advantage regression on two, with no detected increase in scRMSD. The learned biases are also reusable: coefficients fitted on other antigens improve new targets, and AbMPNN-derived biases transfer reward gains to ESM-IF and AntiFold without refitting. Together, these results suggest that direct residue steering can capture transferable design preferences that the tested full-parameter updates struggle to learn.

Limitations and outlook. Our results rely on computational interface metrics and require experimental binding validation. Future work can extend PACT to richer position- or context-dependent biases while preserving its simple fitting and transfer properties (Appendix F.10).

Ethics Statement

This work aims to support antibody discovery through computational sequence design. All results are computational; binding, specificity, and developability require experimental validation. The interface reward does not assess off-target binding or immunogenicity, and aggressive steering can favor larger interfaces without better packing (Appendix F.3). Generated sequences should therefore be treated as candidates for experimental testing. As with other protein-design tools, potential dual-use risks should be considered when releasing or applying the method.

Reproducibility Statement

Algorithm 1 specifies PACT, while Appendix D documents data generation, backbone splits, decoding, rewards, evaluation, and statistical testing. Derivations and gradient diagnostics are given in Appendices A, B and E. We report the complete fitted coefficients in Table 17, the bias-strength sweep in Appendix F.3, and the tested full-parameter objectives and training diagnostics in Appendices D.8 and E.1.

AI Use Statement

We used AI tools to assist with coding, running computational experiments, checking mathematical derivations and statistical results, revising figures and manuscript organization, checking references, and improving the clarity and language of the manuscript. The authors take full responsibility for the implementation, experimental results, scientific claims, and final text.

References

Christof Angermueller, David Dohan, David Belanger, Ramya Deshpande, Kevin Murphy, and Lucy Colwell. Model-based reinforcement learning for biological sequence design. In *International conference on learning representations*, 2020. URL https://openreview.net/forum?id=HklxbgBKvr.

Nathaniel R Bennett, Joseph L Watson, Robert J Ragotte, Andrew J Borst, Déjenaé L See, Connor Weidle, Riti Biswas, Ellen L Shrock, Philip JY Leung, Buwei Huang, et al. Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*, 649(8095):183–193, 2025. doi: 10.1038/s41586-025-09721-5.

Chai Discovery Team, Jacques Boitreaud, Jack Dent, Danny Geisz, Matthew McPartlon, Joshua Meier, Zhuoran Qiao, Alex Rogozhnikov, Nathan Rollins, Paul Wollenhaupt, et al. Zero-shot antibody design in a 24-well plate. *bioRxiv*, 2025. doi: 10.1101/2025.07.05.663018.

Ziqi Chen, Baoyi Zhang, Hongyu Guo, Prashant Emani, Trevor Clancy, Chongming Jiang, Mark Gerstein, Xia Ning, Chao Cheng, and Martin Renqiang Min. Binding peptide generation for MHC Class I proteins with deep reinforcement learning. *Bioinformatics*, 39(2):btad055, 2023.

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*, 2022.

Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. RAFT: Reward ranked finetuning for generative foundation model alignment. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=m7p5O7zblY.

Frédéric A Dreyer, Daniel Cutting, Constantin Schneider, Henry Kenlay, and Charlotte M Deane. Inverse folding for antibody sequence design using deep learning. *arXiv preprint arXiv:2310.19513*, 2023.

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.

Matthew C. Franklin, Kimberly D. Carey, Felix F. Vajdos, Daniel J. Leahy, Abraham M. de Vos, and Mark X. Sliwkowski. Insights into ErbB signaling from the structure of the ErbB2–pertuzumab complex. *Cancer Cell*, 5:317–328, 2004. doi: 10.1016/S1535-6108(04)00083-2.

Johannes Thorling Hadsund, Tadeusz Satława, Bartosz Janusz, Lu Shan, Li Zhou, Richard Röttger, and Konrad Krawczyk. nanoBERT: a deep learning model for gene agnostic navigation of the nanobody mutational space. *Bioinformatics Advances*, 2024.

Brian L Hie, Varun R Shanker, Duo Xu, Theodora UJ Bruun, Payton A Weidenbacher, Shaogeng Tang, Wesley Wu, John E Pak, and Peter S Kim. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 2024.

J. A. Hixon et al. New anti-IL-7R alpha monoclonal antibodies show efficacy against T cell acute lymphoblastic leukemia in pre-clinical models. *Leukemia*, 34:35–49, 2020. doi: 10.1038/s41375-019-0531-8.

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. doi: 10.1093/bioadv/vbae202.

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.

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. In *International Conference on Learning Representations*, 2022. URL https://people.csail.mit.edu/tommi/papers/Jin_etal_ICLR2022.pdf.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3D equivariant graph translation. In *The Eleventh International Conference on Learning Representations*, 2023a. URL https://openreview.net/forum?id=LFHFQbjxIiP.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 17409–17429. PMLR, 2023b. URL https://proceedings.mlr.press/v202/kong23c.html.

Chak Shing Lee, Conor F Hayes, Denis Vashchenko, and Mikel Landajuela. Reinforcement learning for antibody sequence infilling. *bioRxiv*, 2025. doi: 10.1101/2025.08.08.669419.

J. Y. Lee et al. Structural basis of checkpoint blockade by monoclonal antibodies in cancer immunotherapy. *Nature Communications*, 7:13354, 2016. doi: 10.1038/ncomms13354.

Jinwoo Leem, Laura S Mitchell, James HR Farmery, Justin Barton, and Jacob D Galson. Deciphering the language of antibodies using self-supervised learning. *Patterns*, 2022.

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.

Kenneth Murphy and Casey Weaver. *Janeway's immunobiology*. Garland science, 2016.

Tobias H. Olsen, Iain H. Moal, and Charlotte M. Deane. AbLang: an antibody language model for completing antibody sequences. *Bioinformatics Advances*, 2(1):vbac046, 2022. doi: 10.1093/bioadv/vbac046.

Ryan Park, Darren J Hsu, C Brian Roland, Maria Korshunova, Chen Tessler, Shie Mannor, Olivia Viessmann, and Bruno Trentini. Improving inverse folding for peptide design with diversity-regularized direct preference optimization. *arXiv preprint arXiv:2410.19471*, 2024.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707.

Xue Bin Peng, Aviral Kumar, Grace Zhang, and Sergey Levine. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. *arXiv preprint arXiv:1910.00177*, 2019.

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

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*, 2023.

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. doi: 10.1016/j.cels.2023.10.001.

Chenyu Wang, Masatoshi Uehara, Yichun He, Amy Wang, Avantika Lal, Tommi Jaakkola, Sergey Levine, Aviv Regev, Hanchen Wang, and Tommaso Biancalani. Fine-tuning discrete diffusion models via reward optimization with applications to DNA and protein design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/771e09dd204ea339da0d8114c48afd21-Abstract-Conference.html.

Ziwen Wang, Jiajun Fan, Ruihan Guo, Thao Nguyen, Heng Ji, and Ge Liu. ProteinZero: Self-improving protein generation via online reinforcement learning. *Transactions on Machine Learning Research*, 2026. URL https://openreview.net/forum?id=pVvm0CQAr6.

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*, 2023.

Xiaopeng Xu, Tiantian Xu, Juexiao Zhou, Xingyu Liao, Ruochi Zhang, Yu Wang, Lu Zhang, and Xin Gao. Ab-gen: antibody library design with generative pre-trained transformer and deep reinforcement learning. *Genomics, Proteomics & Bioinformatics*, 21(5):1043–1053, 2023.

Fanglei Xue, Andrew Kubaney, Zhichun Guo, Joseph K Min, Ge Liu, Yi Yang, and David Baker. Improving protein sequence design through designability preference optimization. *arXiv preprint arXiv:2506.00297*, 2025.

Xiangxin Zhou, Dongyu Xue, Ruizhe Chen, Zaixiang Zheng, Liang Wang, and Quanquan Gu. Antigen-specific antibody design via direct energy-based preference optimization. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/daef77101ba5711084a57442c8cf2709-Abstract-Conference.html.

APPENDIX CONTENTS

A Optimal Policy and Value Function **14**

B Gradient Geometry and Update Bounds **14**

B.1 Residual decomposition at an arbitrary iterate 14

B.2 A local bound for any optimizer and a conditional path bound 15

C PACT Algorithm and Composition-Tilt Derivations **16**

C.1 From Additive Rewards to Autoregressive Steering 17

C.2 Temperature, Bias Strength, and Shift Invariance 18

C.3 Deriving and Interpreting the Coefficient Estimator 18

D Experimental Setup and Evaluation **19**

D.1 Target Antigens 19

D.2 Data Generation 19

D.3 Fitting the Surrogate Reward 20

D.4 Fitting Setup 21

D.5 Comparison Design 21

D.6 Evaluation Metrics 23

D.7 Computational Cost 24

D.8 Full-Parameter Training Objectives 24

D.9 Checkpoint Probe Protocol 25

E Full-Parameter Training Diagnostics **26**

E.1 Six-Objective Sweep and Refolded Comparisons 26

E.2 Original initialization measurements 28

E.3 Validating the initial response with real rewards 28

E.4 Checkpoint geometry and reward fitting 29

E.5 Local predictions at later checkpoints 31

F Additional Experimental Results **32**

F.1 Detailed Primary and Method Comparisons 32

F.2 Fitted Composition Coefficients 35

F.3 Bias-Strength Sensitivity in an Earlier Pilot 36

F.4 Richer Restricted Families 37

F.5 Choosing the Steering Vector 37

F.6 Archived Steering Diagnostics 37

F.7 Magnitude-Matched Steering Controls 38

F.8 Transfer to Other Decoders 39

F.9 Refolding Oracle Comparison 39

F.10 Scope and Limitations 40

A Optimal Policy and Value Function

For a fixed context C, let $\mathcal{D}(C)$ be the finite support of the normalized reference distribution, with finite rewards on this set. We optimize over all conditional distributions on this set, with $\beta > 0$. This unrestricted optimum need not be attainable by a particular neural-network parameterization.

The per-context objective is

$$J_C(\pi) = \mathbb{E}_{\hat{S}\sim\pi(\cdot|C)}[r(C,\hat{S})] - \beta\,\mathrm{KL}(\pi(\cdot \mid C) \,\|\, \pi_{\mathrm{ref}}(\cdot \mid C))\,. \tag{10}$$

Define the finite, positive normalizer

$$Z(C) = \sum_{\hat{S}\in\mathcal{D}(C)} \pi_{\mathrm{ref}}(\hat{S} \mid C) e^{r(C,\hat{S})/\beta}$$

and the corresponding normalized distribution

$$\pi^*(\hat{S} \mid C) = \frac{\pi_{\mathrm{ref}}(\hat{S} \mid C) e^{r(C,\hat{S})/\beta}}{Z(C)}. \tag{11}$$

Substituting this expression into the KL divergence gives

$$\beta\,\mathrm{KL}(\pi\|\pi^*) = \beta\,\mathbb{E}_\pi\left[\log \frac{\pi(\hat{S} \mid C)}{\pi_{\mathrm{ref}}(\hat{S} \mid C)} - \frac{r(C,\hat{S})}{\beta} + \log Z(C)\right] \tag{12}$$

$$= \beta \log Z(C) - J_C(\pi). \tag{13}$$

The KL divergence is nonnegative and vanishes only at $\pi = \pi^*$. Hence

$$V^*(C) = \max_\pi J_C(\pi) = \beta \log Z(C) = \beta \log \mathbb{E}_{\hat{S}\sim\pi_{\mathrm{ref}}(\cdot|C)}\left[e^{r(C,\hat{S})/\beta}\right]. \tag{14}$$

Since unrestricted conditional distributions can be chosen separately for each context, this also maximizes the context-averaged objective in the main text. Finally, taking the log-ratio in Equation (11) yields $\beta \log(\pi^*/\pi_{\mathrm{ref}}) = r - V^*(C)$, the identity used to motivate advantage regression in Equation (4).

The role of the baseline. A context-shared baseline preserves pairwise reward differences, but its value matters for normalization. In particular,

$$\mathbb{E}_{\pi_{\mathrm{ref}}}\left[e^{(r(C,\hat{S})-\hat{V}(C))/\beta} \mid C\right] = e^{(V^*(C)-\hat{V}(C))/\beta}.$$

Thus, zero loss on the full reference support in Equation (4) is possible only when $\hat{V}(C) = V^*(C)$. By Jensen's inequality, the population mean reward is at most $V^*(C)$, with equality only for a constant reward. A sample mean centers the observed targets; it does not generally supply the normalizer required by the population optimum.

B Gradient Geometry and Update Bounds

We first derive the relative update at an arbitrary iterate, then give a local bound that applies to any optimizer. The original initialization diagnostics are in Appendix E.2; the subsequent checkpoint study uses the sample, training runs, and measurement controls specified in Appendix D.9.

B.1 Residual decomposition at an arbitrary iterate

Fix a context C, n candidate designs, their rewards, a decoding order, a reference policy, and $\beta > 0$. Set $A_i = r(C,\hat{S}_i) - \hat{V}(C)$. Let J_t have rows $g_{t,i}^\top = \nabla_\theta \log \pi_{\theta_t}(\hat{S}_i \mid C)^\top$ and $K_t = J_t J_t^\top$. Vector norms are Euclidean, and $\|\cdot\|_2$ denotes the matrix spectral norm. Write

$$\ell_{t,i} = \beta \log \frac{\pi_{\theta_t}(\hat{S}_i \mid C)}{\pi_{\mathrm{ref}}(\hat{S}_i \mid C)}, \qquad e_t = A - \ell_t, \qquad P = I - \frac{\mathbf{1}\mathbf{1}^\top}{n}, \qquad Q = I - P.$$

For the empirical mean loss $\mathcal{L}_C = n^{-1}\|\ell_t - A\|^2$, a gradient step of size h_t has parameter increment

$$\Delta\theta_t = \frac{2h_t\beta}{n} J_t^\top e_t. \tag{15}$$

The Jacobian of ℓ is βJ_t, so Taylor expansion and $P + Q = I$ give

$$\Delta\ell_t = \eta_t K_t e_t + R_t, \qquad \eta_t = 2h_t\beta^2/n, \tag{16}$$

$$P\Delta\ell_t = \eta_t PK_tP\, Pe_t + \eta_t PK_tQ\, Qe_t + PR_t. \tag{17}$$

Here R_t is the remainder of the real-valued scoring map; with bounded second derivatives, $\|R_t\| = O(\|\Delta\theta_t\|^2)$. Finite-precision evaluation adds numerical error to the measured discrepancy.

The first term responds to differences in the residual across candidates. The second shows how its mean couples into those relative updates. For the mean-reward baseline, $QA = 0$ and $Qe_t = -\bar{\ell}_t\mathbf{1}$, yielding Equation (5). This coupling vanishes at reference initialization, when $\ell_0 = 0$. We then recover

$$P\Delta\ell_0 = \eta_0 PK_0PA + PR_0, \qquad \|P\Delta\ell_0\| \le \eta_0\|PK_0P\|_2\|A\| + \|PR_0\|. \tag{18}$$

For a soft-value baseline, QA need not vanish, so the initial first-order response is $\eta_0 PK_0A$. At later iterates, replacing K_0 with K_t alone misses both the changing residual and its mean-coupling term.

What centering removes. Alignment with the constant direction matters more than rank alone. If $K = uu^\top$ with $u = c\mathbf{1}$, then $PKP = 0$; a nonconstant u can yield $PKP \neq 0$ even at rank one. The ratio $\|PKP\|_2/\|K\|_2$ compares centered and uncentered norm bounds. A small ratio alone establishes neither slow optimization in absolute units nor poor reward correlation: for $PA \neq 0$, even an arbitrarily small positive multiple of PA has correlation one with PA.

Clipping and joint updates. For an SGD step on this fixed-context loss, scalar gradient clipping multiplies both first-order response terms by the clipping factor. An update using other contexts also introduces cross-context blocks $J_C J_{C'}^\top$ and their sample weights. AdamW additionally depends on accumulated moments, preconditioning, and weight decay, so its direction is not generally determined by the current context's raw kernel. The following bound instead uses the actual parameter increment.

B.2 A LOCAL BOUND FOR ANY OPTIMIZER AND A CONDITIONAL PATH BOUND

Let $d_t = \theta_{t+1} - \theta_t$ be any actual optimizer increment. Independently of how it was computed,

$$P\Delta\ell_t = \beta PJ_td_t + PR_t, \qquad \|\beta PJ_td_t\| \le \beta\sqrt{\|PK_tP\|_2}\,\|d_t\|. \tag{19}$$

This follows from $\|PJ_t\|_2^2 = \|PJ_tJ_t^\top P\|_2 = \|PK_tP\|_2$. Measured increments include momentum, preconditioning, decay, and parameter rounding. The bound applies to the linear prediction; the realized update also includes the remainder. Our probes compare both direction and magnitude of this prediction with the observed change.

An exact path statement is also possible. Suppose the scoring map is differentiable and its restriction to each parameter segment is absolutely continuous. Then

$$P(\ell_{t+1} - \ell_t) = \beta\int_0^1 PJ(\theta_t + sd_t)d_t\, ds. \tag{20}$$

Summing over every actual training step and applying the triangle inequality gives

$$\|P(\ell_T - \ell_0)\| \le \beta\sum_{t<T}\|d_t\|\int_0^1\sqrt{\|PK(\theta_t + sd_t)P\|_2}\, ds \le \beta\rho L_T, \tag{21}$$

where $L_T = \sum_{t<T}\|d_t\|$ and ρ bounds the square-root centered-kernel norm over the whole path, including each segment. Starting from the reference, $\ell_0 = 0$, so

$$\|P(A - \ell_T)\| \ge \|PA\| - \beta\rho L_T. \tag{22}$$

This lower bound is informative only when its right-hand side is positive. It does not require a constant kernel or a particular optimizer, but does require bounds on ρ and L_T. Sparse checkpoints do not bound the kernel between saved iterates, and the sum of checkpoint-to-checkpoint distances is a lower bound on the historical path length, not an upper bound. We therefore do not claim a nonvacuous full-trajectory bound for these runs. The checkpoint evidence in Appendices E.4 and E.5 establishes persistence of measured structure and accuracy of local predictions, rather than a causal explanation of long-run training failure.

C PACT ALGORITHM AND COMPOSITION-TILT DERIVATIONS

Algorithm 1 summarizes the fitting and decoding procedure. We detail each stage below.

Algorithm 1 PACT

Require: fitting contexts D_{fit}; frozen reference decoder with logits z; budget N; tilt $1/\beta$; temperature τ; fitting pipeline $\mathcal{O}_{\text{fit}}$
Require: deployment context C_{new}
$\mathcal{R} \leftarrow \emptyset$
for $C \in D_{\text{fit}}$ **do**
  sample $\{\hat{S}_i\}_{i=1}^{N}$ from the reference at temperature τ, with a random decoding order
  $r_i \leftarrow \textsc{Score}(\mathcal{O}_{\text{fit}}(\hat{S}_i, C))$ for each i ▷ refold, then Rosetta interface metrics
  $\bar{r}(C) \leftarrow \frac{1}{N}\sum_i r_i$; $\mathcal{V}(C) \leftarrow \{q \in \mathcal{I}_{\text{CDR}} : |\{\hat{S}_{i,q}\}_i| > 1\}$
  for $i = 1..N$, $q \in \mathcal{V}(C)$ **do** ▷ one entry per *occurrence*, not per sequence
    append $\big(\hat{S}_{i,q},\ r_i - \bar{r}(C)\big)$ to the multiset $\mathcal{R}$
  end for
end for
for each $a \in \mathcal{A}$ **do**
  $\hat{b}_a \leftarrow \text{mean}\{\,\delta : (a, \delta) \in \mathcal{R}\,\}$ if a occurs ≥ 150 times, else 0 ▷ Equation (9)
end for
$C \leftarrow C_{\text{new}}$; initialize the fixed residues from C
sample a random decoding order of the designable CDR positions
for each designable position q in that order **do**
  $\pi^{\text{loc}}_{\hat{b}}(\cdot \mid \hat{S}_{<q}, C) \leftarrow \text{softmax}\big(\frac{1}{\tau}[\,z(\cdot \mid \hat{S}_{<q}, C) + \hat{b}/\beta\,]\big)$ ▷ Equation (8)
  sample $\hat{S}_q \sim \pi^{\text{loc}}_{\hat{b}}(\cdot \mid \hat{S}_{<q}, C)$
end for
return $\hat{b}$, $\hat{S}$ ▷ reuse $\hat{b}$ for further designs; network weights stay fixed

Inputs. A set of fitting contexts D_{fit} (each an antigen/backbone prompt), a pretrained reference policy π_{ref}, a fitting designs-per-context budget N, the tilt magnitude $1/\beta$, the temperature τ, deployment contexts, and two refolding oracles: $\mathcal{O}_{\text{fit}}$ (IgFold followed by framework alignment to the designed pose) and $\mathcal{O}_{\text{eval}}$ (complex re-prediction; RF2).

Stage 1 — Sampling. Contexts are split disjointly into fitting and evaluation sets. For each fitting context C we draw N designs from the reference at low temperature; diversity arises mainly from the random decoding order of the autoregressive sequence model, so designs can differ even when each conditional choice is effectively greedy. In Algorithm 1, $\hat{S}_{<q}$ denotes the residues preceding position q in the sampled order, not in sequence index order.

Stage 2 — Reward evaluation under the fitting oracle. Each design is refolded with $\mathcal{O}_{\text{fit}}$ and scored with Rosetta's `InterfaceAnalyzer`, giving $r(C, \hat{S}_i)$. The fitting pipeline matters because Equation (9) is an average over these observations, so oracle sampling noise propagates directly into the estimated coefficients (Section 4 and Appendix F.9).

Stage 3 — Closed-form coefficient estimation. Within each context we identify the positions $\mathcal{V}(C)$ that vary across the N designs and center the rewards by the within-context mean $\bar{r}(C)$, which removes context-level offsets that carry no information about which design is better. Each residue's coefficient $\hat{b}_a$ is then the mean centered reward over all *occurrences* of a at those positions — a design contributing a at three variable positions contributes three observations, not one (Equation (9)). A common shift of all coefficients leaves every per-position softmax unchanged, leaving 19 effective degrees of freedom in the unrestricted bias family. We use the fitted vector directly. Subtracting its mean would preserve the policy but change its L_1 norm, so the magnitude-matched controls also use the raw-vector norm (Appendix C.2). This is a single pass over the data; there is no optimization loop, learning rate, or convergence criterion.

Stage 4 — Deployment and evaluation. At inference, $\hat{b}/\beta$ is added to the decoder logits at each designable position, inside the per-position softmax and before temperature scaling, leaving all network weights untouched. The deployed policy is therefore the locally renormalized Equation (8) rather than the sequence-level tilt of Equation (7). Evaluation decodes from the tilted policy on the held-out contexts, refolds with $\mathcal{O}_{\text{eval}}$, and compares against π_{ref}; differences are averaged within each backbone and tested across backbones, since designs sharing a backbone are not independent. We report the aggregate reward together with its components, since a composition change can raise the aggregate by enlarging the interface without improving its quality.

C.1 From Additive Rewards to Autoregressive Steering

Fix a context C and one decoding order. In this subsection, $t = 1, \dots, m$ indexes the designable positions in that order, and $s_{<t}$ denotes the residues already decoded. All distributions below are conditional on this fixed order, and π_{ref} denotes the unit-temperature reference distribution in this derivation. Let $u_b(s) = \sum_{t=1}^{m} b_{s_t}$ be the composition contribution to the reward.

The sequence-level distribution. For the additive reward model $r_b(C, s) = r_0(C) + u_b(s)$, the Boltzmann normalizer factors as

$$\mathbb{E}_{\pi_{\text{ref}}}[e^{r_b(C,s)/\beta} \mid C] = e^{r_0(C)/\beta}\, \mathcal{Z}_b(C), \qquad \mathcal{Z}_b(C) = \mathbb{E}_{\pi_{\text{ref}}}[e^{u_b(s)/\beta} \mid C]. \tag{23}$$

The factor $e^{r_0(C)/\beta}$ cancels from the numerator and denominator, giving

$$\pi_b(s \mid C) = \frac{\pi_{\text{ref}}(s \mid C) e^{u_b(s)/\beta}}{\mathcal{Z}_b(C)}. \tag{24}$$

This is exact for r_b. When r_b only approximates the measured reward, it motivates the composition family in Equation (7) without identifying the optimum for that measured reward.

The locally normalized policy. Write the reference conditional at temperature $\tau > 0$ as

$$p_t^{(\tau)}(a \mid s_{<t}, C) = \frac{e^{z_{t,a}(s_{<t},C)/\tau}}{\sum_{a'} e^{z_{t,a'}(s_{<t},C)/\tau}}.$$

Adding b/β before temperature scaling gives

$$\pi_b^{\text{loc}}(a \mid s_{<t}, C) = \frac{e^{z_{t,a}/\tau} e^{b_a/(\beta\tau)}}{\sum_{a'} e^{z_{t,a'}/\tau} e^{b_{a'}/(\beta\tau)}} = \frac{p_t^{(\tau)}(a \mid s_{<t}, C) e^{b_a/(\beta\tau)}}{Z_t(s_{<t}, C)}, \tag{25}$$

$$Z_t(s_{<t}, C) = \sum_{a'} p_t^{(\tau)}(a' \mid s_{<t}, C) e^{b_{a'}/(\beta\tau)}. \tag{26}$$

Here the logit arguments are suppressed in the first fraction. Multiplying the conditionals along the decoding order yields

$$\frac{\pi_b^{\text{loc}}(s \mid C)}{\pi_{\text{ref}}^{(\tau)}(s \mid C)} = \frac{\exp(u_b(s)/(\beta\tau))}{\prod_{t=1}^{m} Z_t(s_{<t}, C)}, \qquad \pi_{\text{ref}}^{(\tau)}(s \mid C) = \prod_{t=1}^{m} p_t^{(\tau)}(s_t \mid s_{<t}, C). \tag{27}$$

At $\tau = 1$, the numerator matches the sequence-level tilt in Equation (24). The policies coincide exactly when the product of local normalizers is independent of s on the reference support; normalization then forces that product to equal $\mathcal{Z}_b(C)$. Prefix-independent local normalizers are sufficient for this equality. An autoregressive decoder generally has prefix-dependent logits, so the equality is not guaranteed.

The difference can also be seen in the exact conditional distribution of the sequence-level tilt. At $\tau = 1$, define the expected contribution of the remaining positions as

$$F_t(s_{\le t}, C) = \mathbb{E}_{\pi_{\text{ref}}(\cdot \mid s_{\le t}, C)}\left[\exp\left(\frac{1}{\beta} \sum_{u=t+1}^{m} b_{s_u}\right)\right], \qquad F_m = 1.$$

Marginalizing the future residues in Equation (24) gives

$$\pi_b(a \mid s_{<t}, C) = \frac{p_t^{(1)}(a \mid s_{<t}, C) e^{b_a/\beta} F_t(s_{<t}a, C)}{\sum_{a'} p_t^{(1)}(a' \mid s_{<t}, C) e^{b_{a'}/\beta} F_t(s_{<t}a', C)}. \tag{28}$$

The future factor F_t can depend on the current residue through the reference decoder. PACT applies the local bias without computing this factor. It is therefore defined by Equation (8), with Equation (7) providing its motivation. The experiments randomize the decoding order for each design; the identities above apply conditional on each order.

C.2 Temperature, Bias Strength, and Shift Invariance

For two amino acids a and a' at the same decoded prefix, the local log-odds satisfy

$$\log \frac{\pi_b^{\mathrm{loc}}(a \mid s_{<t}, C)}{\pi_b^{\mathrm{loc}}(a' \mid s_{<t}, C)} = \frac{z_{t,a} - z_{t,a'}}{\tau} + \frac{b_a - b_{a'}}{\beta\tau}. \tag{29}$$

Thus τ scales both the reference preference and the added bias, while $1/\beta$ changes only the bias. As $\tau \to 0^+$, probability concentrates on maximizers of $z_{t,a} + b_a/\beta$; the maximizer itself is independent of τ. Our temperature $\tau = 10^{-4}$ makes each conditional effectively greedy, with diversity arising mainly from the random decoding order.

For any scalar c, adding $c\mathbf{1}$ to b adds the same $c/(\beta\tau)$ to every softmax input, which cancels in normalization:

$$\pi_{b+c\mathbf{1}}^{\mathrm{loc}}(a \mid s_{<t}, C) = \pi_b^{\mathrm{loc}}(a \mid s_{<t}, C).$$

Conversely, with finite logits and fixed positive β, τ, equality of all conditional probabilities implies equality of all log-odds, so two bias vectors differ only by a constant. The unrestricted 20-entry bias family therefore has $20 - 1 = 19$ effective degrees of freedom. The same invariance holds for Equation (7), since $u_{b+c\mathbf{1}}(s) = u_b(s) + mc$ and m is fixed within a context. The frequency threshold used for fitting fixes two or three entries at zero, leaving 17–18 estimated coefficients within this family.

We use $\hat{b}$ directly, without subtracting its mean across the 20 amino-acid types. Subtracting that mean would preserve the policy but can change $\|\hat{b}\|_1$. The magnitude-matched controls use the L_1 norm of the raw fitted vector, so we retain this convention when reporting coefficient norms. Rewards are still centered within each backbone before the coefficients are estimated (Equation (9)).

C.3 Deriving and Interpreting the Coefficient Estimator

For design s_i from context C, let $\delta_{Ci} = r(C, s_i) - \bar{r}(C)$. Form a multiset $\mathcal{N}$ containing one triple (C, i, q) for each position $q \in \mathcal{V}(C)$ of each sampled design. An occurrence-level predictor assigns a scalar u_a to each amino-acid identity and fits the centered sequence reward associated with that occurrence:

$$\mathcal{J}(u) = \sum_{(C,i,q)\in\mathcal{N}} (\delta_{Ci} - u_{s_{i,q}})^2 = \sum_{a\in\mathcal{A}} \sum_{(C,i,q)\in\mathcal{N}_a} (\delta_{Ci} - u_a)^2. \tag{30}$$

This objective separates over amino acids. For $N_a = |\mathcal{N}_a| > 0$,

$$\frac{\partial \mathcal{J}}{\partial u_a} = 2\left(N_a u_a - \sum_{(C,i,q)\in\mathcal{N}_a} \delta_{Ci}\right) = 0 \quad\Longrightarrow\quad \hat{b}_a = \frac{1}{N_a} \sum_{(C,i,q)\in\mathcal{N}_a} \delta_{Ci}.$$

This is the unthresholded estimator in Equation (9). PACT additionally sets coefficients with fewer than 150 occurrences to zero. The resulting vector minimizes $\mathcal{J}$ subject to $u_a = 0$ for these residues; the threshold itself is an estimation rule, not a consequence of unconstrained minimization. Cysteine and histidine remain zero on all three antigens; tryptophan also remains zero on IL7R and PD1 (Table 17).

The fitted quantity is a mean over residue occurrences. It differs from fitting the sum of residue contributions to each sequence's reward. To make this distinction explicit, define

$$m_{Ci,a} = \sum_{q\in\mathcal{V}(C)} \mathbf{1}\{s_{i,q} = a\}, \qquad \bar{m}_{C,a} = \frac{1}{n_C} \sum_i m_{Ci,a}, \qquad N_a = \sum_{C,i} m_{Ci,a},$$

where n_C is the number of sampled designs for C. For a residue retained by the frequency threshold, $\sum_i \delta_{Ci} = 0$ gives

$$\hat{b}_a = \frac{\sum_{C,i} m_{Ci,a}\delta_{Ci}}{N_a} = \frac{\sum_{C,i} (m_{Ci,a} - \bar{m}_{C,a})\delta_{Ci}}{N_a}. \tag{31}$$

Even if the sampled rewards follow an exactly additive model $r(C, s_i) = \alpha_C + \sum_c \gamma_c m_{Ci,c}$, subtracting the context mean gives $\delta_{Ci} = \sum_c \gamma_c (m_{Ci,c} - \bar{m}_{C,c})$. Substitution yields

$$\hat{b}_a = \frac{1}{N_a} \sum_c \gamma_c \sum_{C,i} (m_{Ci,a} - \bar{m}_{C,a})(m_{Ci,c} - \bar{m}_{C,c}). \tag{32}$$

The estimate therefore depends on count variation and co-occurrence across residue types, and is not generally equal to the additive coefficient γ_a. The retained coefficients minimize the occurrence-level prediction objective in Equation (30), but provide heuristic steering coefficients rather than an exact fit of the additive sequence reward. Its derivation also does not establish optimality for the original KL-regularized objective.

The algebra does not require independent occurrences. Repeated use of a sequence's reward does, however, make those observations dependent. We therefore assess coefficient stability by splitting backbones, as reported in Appendix F.2, rather than treating the occurrence count as an effective sample size.

D Experimental Setup and Evaluation

The setup follows the data and scoring pipeline, then specifies comparison protocols, computational cost, and training diagnostics. Supplementary optimization results are in Appendix E; design results and visualizations are in Appendix F.

D.1 Target Antigens

We use epitopes from three experimentally determined antibody–antigen complexes. The chain IDs and residue numbers below follow the PDB author numbering, before conversion to `HLT` format.

- **HER2.** PDB 1S78 contains the HER2–pertuzumab Fab complex (Franklin et al., 2004). We target the pertuzumab epitope on HER2 domain II, using chain A hotspots A286, A290, A295, and A296. Our modeled HER2 complex has 675 residues, the largest complex in this study; this is a sequence-length count, not an interface-area measure.
- **IL7R.** PDB 6P50 contains human IL-7 receptor α bound to the 4A10 Fab (Hixon et al., 2020). We use chain C hotspots C52, C61, C84, C86, C88, C90, C95, and C100. The target is the receptor ectodomain, rather than the IL-7 cytokine.
- **PD1.** PDB 5GGS contains PD1 bound to pembrolizumab Fab (Lee et al., 2016). We use chain Y hotspots Y62, Y63, Y66, Y78, Y85, Y87, Y128, and Y131. PD1 has the smallest modeled complex in this study. Repeated PD1 predictions provide the oracle diagnostics in Appendix F.9.

D.2 Data Generation

We use the backbone-generation and sequence-design pipeline of Bennett et al. (2025), with separate refolding procedures for coefficient fitting and held-out evaluation.

Input representation. Each context is stored as an annotated PDB file in `HLT` format: chains are ordered as heavy (`H`), light (`L`, if present), and target (`T`), with `REMARK` records identifying CDR residues. These annotations define the framework and designable positions consistently across generation, decoding, and scoring. Chothia-annotated structures from SAbDab (Dunbar et al., 2014) can be converted to this format.

Pipeline stages. Each antigen is processed in four sequential steps:

1. **RFdiffusion:** An antibody-fine-tuned diffusion model generates backbone structures against specified epitopes. CDRs are flexibly designed within user-defined length ranges, while non-loop framework regions remain fixed.
2. **AbMPNN:** Designed backbones are converted to sequences by assigning CDR residues. This ensures sequence variability while preserving framework integrity.

3. **Refolding:** IgFold predicts fitting structures, followed by framework alignment to the designed pose. An antibody-fine-tuned RoseTTAFold2 (RF2) model predicts evaluation complexes, allowing us to assess structural self-consistency and interface geometry.

4. **Rosetta analysis:** Refolded complexes are evaluated with Rosetta's `InterfaceAnalyzer`. Three interface metrics, shape complementarity (SC), buried solvent-accessible surface area (ΔSASA), and the number of cross-interface hydrogen bonds (Hbonds), are extracted. These are combined into the surrogate reward used for steering and evaluation.

Backbone-generation settings. Antibody-fine-tuned RFdiffusion (Bennett et al., 2025; Watson et al., 2023) allows CDR-H1 lengths of 6–10 residues, H2 lengths of 7–11, and H3 lengths of 5–20. The configuration is `antibody.design_loops=[H1:6-10,H2:7-11,H3:5-20]`. The recorded inference settings are `diffuser.T=50`, `inference.final_step=48`, and `inference.deterministic=True`. The last two values specify the stopping step and enable the implementation's deterministic inference option.

Backbone split. For each antigen, we generate 256 backbones and reserve IDs 0–127 for evaluation. IDs 128–255 form the coefficient-fitting pool. Full-parameter training uses this same fitting pool, with an internal validation split as specified in Appendices D.8 and D.9. No evaluation backbone is used to estimate the deployed coefficient vectors.

D.3 Fitting the Surrogate Reward

We estimate the reward weights on the antibody–antigen subset of SKEMPI v2.0 (Jankauskaitė et al., 2019). The regression predicts experimental $\Delta\Delta G$ from mutant–wild-type feature differences, $\delta m = m_{\mathrm{mut}} - m_{\mathrm{wt}}$. Of the four initial predictors (Table 3), the energy-density term $100\,\Delta G_{\mathrm{sep}}/\Delta\mathrm{SASA}$ is not significant ($t = -0.44$, $p = 0.66$). We drop it and refit on the remaining three terms (Table 4). Variance-inflation factors are close to one, indicating little multicollinearity in this fit.

Table 3: Full regression of experimental $\Delta\Delta G$ on mutant–wild-type changes in four Rosetta interface terms (SKEMPI v2.0, antibody–antigen subset).

Metric	Coefficient	Std. error	t	p	VIF
$\Delta G_{\mathrm{sep}}/\Delta\mathrm{SASA}\times 100$	-0.895	2.014	-0.44	0.657	1.031
ΔSASA (normalized)	-38.456	2.906	-13.23	2.1×10^{-39}	1.140
Hbonds (normalized)	-10.969	1.979	-5.54	3.1×10^{-8}	1.125
SC	-12.949	0.973	-13.30	8.3×10^{-40}	1.014

Table 4: Refit on mutant–wild-type feature differences after dropping the energy-density term. Absolute coefficient values give the reward weights $|w_i|$.

Metric	Coefficient	Std. error	t	p	VIF
ΔSASA (normalized)	-38.261	2.873	-13.32	6.9×10^{-40}	1.108
Hbonds (normalized)	-11.052	1.970	-5.61	2.1×10^{-8}	1.117
SC (raw)	-12.969	0.972	-13.34	5.2×10^{-40}	1.013

All fitted coefficients are negative, associating increases in the features from wild type to mutant with lower $\Delta\Delta G$ in this regression; we use their absolute values when constructing the reward so that higher values receive higher reward. The score has no intercept, uses fixed divisors of 9200 for ΔSASA and 394 for Hbonds, and is not rescaled per target. Coefficient standard errors are approximately 1–3 (Table 4); Appendix D.6 gives the weights at the precision used in computation. We report no held-out predictive validation of the SKEMPI fit. Applying this fit from mutations in existing complexes to multi-residue CDR redesigns also introduces distribution shift (Appendix F.10).

D.4 Fitting Setup

PACT adds the shared residue bias in Equation (8) to a frozen decoder. We estimate the coefficients in closed form using Equation (9), without gradient training or an optimizer. The unrestricted bias family has 19 effective degrees of freedom; Table 17 gives the complete fitted vectors.

Fitting procedure. For each antigen we (i) decode 16 designs per backbone at temperature 10^{-4} on the 128 *fitting* backbones, (ii) refold each design with IgFold (Ruffolo et al., 2023) and superpose it onto the designed pose through its framework residues, (iii) score with Rosetta's `InterfaceAnalyzer`, and (iv) average the within-context-centered reward per residue over the positions that vary within each context. IgFold refolding costs ∼2.1 s per design and Rosetta interface scoring ∼9.4 s (Table 5), so steps (ii)–(iv) cost ∼6.5 CPU-hours over the 2,048 fitting designs. This excludes GPU sequence decoding and the RF2 scale-selection sweep. The 128 evaluation backbones are never seen during fitting.

Decoding temperature and order. We use $\tau = 10^{-4}$ in Equation (8), so each conditional choice is effectively greedy. Each design uses a random decoding order, which supplies most of the sequence diversity. The bias is added before temperature scaling: $1/\beta$ changes its strength, while τ scales both the reference logits and the bias (Appendix C.2).

Bias strength and frequency threshold. The scalar $1/\beta$ in Equation (8) controls bias strength. In the main 16-design comparison and its eight-design method-comparison subset, we use $1/\beta = \mathrm{round}(1.9/\|\hat{b}\|_\infty)$: the multipliers are 10 for HER2, 23 for IL7R, and 25 for PD1. This makes the largest absolute logit bias approximately 1.9 on each target (1.97, 1.87, and 1.93, respectively). We estimate a coefficient only after observing at least 150 occurrences of that residue at variable positions, setting rarer entries to zero. This frequency threshold was fixed rather than tuned.

Settings for the bias-vector controls. The separate bias-vector comparison uses the same coefficient-estimation procedure but fixes $1/\beta = 19$ for all targets and controls. For the saved vectors in Table 17, this gives total absolute logit biases $\|\hat{b}\|_1/\beta$ of 14.5 on HER2, 9.9 on IL7R, and 10.1 on PD1. The largest individual bias in this comparison is for tryptophan on HER2: $0.197 \times 19 \approx 3.7$ logits. All reported norms and magnitude-matched controls use the raw fitted vectors, without subtracting their mean (Appendix C.2). Comparisons within each experiment use consistent settings; differences between the method and bias-vector comparisons reflect both bias strength and sequence sampling.

Strength selection. The earlier HER2 sweep in Table 18 used a separate fitted vector and a nominal reference scale of 20. It motivated a maximum individual bias near 1.9, but it is not a sweep of the saved IgFold coefficients or a calibration of the controls' common scale of 19. The sweep and subsequent protocol development used RF2 outcomes, so the total model-selection cost exceeds coefficient fitting alone.

Compute. Evaluation dominates the cost. In the reported runs, RF2 refolding takes approximately 30 s per IL7R or PD1 design and 49 s per HER2 design. Evaluating 2,048 designs from one method therefore requires about 17–28 GPU-hours; a paired comparison evaluates both methods. IgFold refolding of the same number of fitting designs takes about 1.2 CPU-hours, plus 5.3 CPU-hours for Rosetta scoring. Table 5 separates these stages.

D.5 Comparison Design

The comparisons separate three questions: how bias steering compares with weight updates, which bias coefficients produce the gain, and whether those coefficients transfer to other generators.

Full-parameter adaptation of AbMPNN. Advantage regression and filtered SFT update all 1.66M weights of the AbMPNN reference that PACT leaves frozen. They use the fitting split and are evaluated on held-out backbones through the same RF2 and Rosetta pipeline. These comparisons isolate the adaptation procedure more closely than comparisons between independently pretrained

generators. Filtered SFT is evaluated after near-exact imitation of high-reward fitting designs; its checkpoint-selection limitation is discussed in Appendix E.1.

Bias-vector controls on frozen AbMPNN. We compare the target-fitted vector with four alternatives. The shuffled vector permutes its coefficients across amino-acid types. The size prior uses side-chain volume to set the direction, while retaining the target vector's support and L_1 norm. The other two vectors are fitted on one other antigen or averaged over the other two antigens. The transferred coefficients use no sequences from the evaluation antigen. All vectors use the common multiplier of 19. A separate diagnostic rescales transferred vectors to the target-fitted norm to distinguish changes in direction from changes in magnitude (Appendix F.7).

Other sequence generators and decoder transfer. ESM-IF (Hsu et al., 2022) and AntiFold (Høie et al., 2025) receive the same backbone and designable positions as AbMPNN (Dreyer et al., 2023). AntiFold uses the ESM-IF GVP-Transformer architecture with antibody-fine-tuned weights. AbLang (Olsen et al., 2022) and nanoBERT (Hadsund et al., 2024) are sequence-only references and receive no antigen context. All generated designs pass through the same refolding and scoring pipeline, but pretraining and conditioning differ across models. Their absolute performance therefore does not isolate the effect of PACT.

For the direct reward comparisons with these four generators, we correct over twelve tests (three targets per generator; $\alpha' = 0.0042$). Results are reported in Appendix F.1.

A separate transfer experiment adds the AbMPNN-derived bias to ESM-IF and AntiFold without refitting. Each steered decoder is compared with its own unsteered output (Figure 6), so these paired comparisons directly test the added bias.

Implementation coverage. We do not tabulate AbDPO (Zhou et al., 2024): an official implementation was unavailable to us, and our reconstruction did not reproduce its full training procedure. Reporting that partial implementation under the method's name would not provide a reliable comparison.

Pairing and inference. Every comparison is paired on the backbone: differences are averaged within each backbone and tested across the available held-out backbones, because designs decoded for one backbone share a structure and are not independent. Main-text comparisons use 128 backbones except HER2–AntiFold transfer, which has 127 (Table 23).

Test families and generation budgets. Primary reward and SC tests each form a three-target family ($\alpha' = 0.0167$). The twelve direct comparisons against alternative generators use $\alpha' = 0.0042$, and the six cross-policy reward-transfer tests use $\alpha' = 0.0083$. The three direct PACT–filtered-SFT tests form their own three-target family, as do the three direct PACT–advantage-regression tests (Table 6). Other reward components, auxiliary metrics, and steering controls are exploratory, including the direct contrasts between steering vectors in Table 20. The main PACT–AbMPNN comparison (Figure 4) uses 16 designs per backbone. The method comparison (Table 1) and bias-vector comparison (Table 2) use eight on the same held-out backbones. Their sequence samples differ, and the bias-vector comparison uses a common multiplier of 19 instead of the method comparison's target-specific strengths (Appendix D.4). The saved refits reproduce the original coefficients. Each comparison uses its corresponding AbMPNN outputs as the baseline. For example, the PACT HER2 reward estimates are $+0.634$, $+0.574$, and $+0.786$ in these three comparisons, respectively. We use backbones as the units of inference throughout, rather than treating decoded designs as independent replicates.

D.5.1 Precision of Shape-Complementarity Comparisons

The main SC effects are approximately 0.02, with standard errors of 0.004–0.005 over 128 backbones. A smaller pilot can give a substantially less precise estimate. We therefore use the full held-out backbone set for the reported primary SC tests and correct over the three targets. Reward and SC are tested separately: an aggregate score gain can be driven by added interface area, so it does not by itself demonstrate improved packing. In the primary comparison, the contribution of SC also varies by target (Figure 4).

D.6 Evaluation Metrics

We report the aggregate reward, its three components, and additional diagnostics of Rosetta separated energy and structural self-consistency. The archived control panel also reports cross-target reference-sequence similarity.

Reward components. We compute SC, ΔSASA, and Hbonds in one Rosetta interface-analysis pass per structure using the REF2015 score function. The `InterfaceAnalyzerMover` enables `pack_input`, `pack_separated`, and `interface_sc`. We use no FastRelax or repeated-repack averaging for these results. Confidence intervals describe variation across backbone-level mean differences; they do not separately estimate the variability of repeated refolding or repacking for an individual design.

ΔG_{sep}: raw and relaxed diagnostics. Raw separated energy is not included in the reward. The candidate energy-density predictor $100\,\Delta G_{\text{sep}}/\Delta\text{SASA}$ was dropped from the SKEMPI regression (Table 3). The energy comparison in Figure 5a uses the raw scorer output on RF2 predictions without FastRelax. These values are sensitive to clashes: mean raw energies are strongly positive (Table 14), so favorable shifts cannot be interpreted as binding free-energy improvements. The separate IL7R control panel in Table 21 instead reports medians after one round of FastRelax; its reported medians are negative. Raw and relaxed results are distinct diagnostics and should not be compared numerically. Both are unoptimized checks from the same Rosetta scoring pipeline used for the reward, rather than independent validation.

The metrics have the following interpretations:

- **Shape complementarity (SC):** A geometric descriptor of how closely the two molecular surfaces match across the interface. Higher values mean tighter steric fit and fewer voids. It is the reward component least directly coupled to interface size, which is why we test it separately.
- **Interface buried area (ΔSASA):** The reduction in solvent-accessible surface area on binding. Larger values mean more surface is buried. Note that this can be increased simply by placing bulkier residues, which is why an aggregate gain driven by ΔSASA alone is not evidence of a better interface.
- **Interface hydrogen bonds (Hbonds):** The number of hydrogen bonds Rosetta identifies across the interface. The reported mean counts are small (approximately 1–3 per design), and individual counts are discrete.
- **Separated energy (ΔG_{sep}):** Rosetta's energy difference between the complex and separated partners, in Rosetta Energy Units (REU). Lower values are favorable under this scoring function. We distinguish the raw and FastRelax-based protocols above; neither is a measured binding free energy.
- **Structural self-consistency (scRMSD):** Heavy-chain C_α root-mean-square deviation (RMSD) between the input scaffold and the RF2 prediction, after Kabsch alignment of the heavy chain. This measures agreement of the antibody backbone, not its pose relative to the antigen. Lower values indicate closer structural agreement. The analysis and coverage are given in Appendix F.1.1.
- **Cross-target reference similarity (Sim):** Positional identity of concatenated CDR sequences to a fixed reference pool from other antigens, using the archived sampling procedure below. This is a descriptive sequence diagnostic, not a test of off-target binding.

These diagnostics describe different properties of the same predicted designs. SC and buried area are geometric quantities; hydrogen bonds and separated energy depend on Rosetta's scoring definitions. They can be correlated and share errors in the predicted structures, so agreement among them does not constitute independent evidence of experimental binding.

Exact reward computation. The displayed equation rounds the weights, while the score calculation uses

$$R = 12.96874668\,\text{SC} + 38.26070987\,\Delta\text{SASA}/9200 + 11.05186575\,\text{Hbonds}/394.$$

There is no intercept or per-target rescaling. These digits specify the implemented score, not the precision of the estimated coefficients: their standard errors are 0.97, 2.87, and 1.97 (Table 4).

All held-out interface comparisons in the main paper use RF2 predictions. IgFold repeatability measurements characterize the fitting pipeline rather than held-out design performance.

Archived Sim aggregation. The archived implementation uses up to 400 alphabetically selected designs from the evaluated method and a fixed pool of up to 200 AbMPNN designs from each of the other two antigens. It samples 20,000 sequence pairs with replacement (seed 0), concatenates the annotated H1–H3 residues, and computes positional identity over the shorter concatenated sequence. No sequence alignment is performed. These values compare each method to a common reference pool; they are not all-pairs similarities between that method's outputs on different targets. The alphabetical selection is not backbone-balanced, so the values in Table 21 remain descriptive. A balanced, loop-aligned cross-target analysis would be needed for a stronger sequence-distinctness claim.

D.7 Computational Cost

Table 5 summarizes the computational budget per antigen: the number of sequences at each stage and the per-structure cost (RFdiffusion backbone generation, RF2 refolding, and Rosetta interface scoring). IgFold provides fitting rewards, while RF2 re-predicts the evaluation complexes (Sections 4 and 5.1). RF2 prediction is distinct from the Rosetta FastRelax step used for the relaxed energy diagnostic. Hardware specifications are listed below.

Table 5: Computation budget per antigen for one fitting pool and one evaluated method. Sequence decoding counts both stages; a paired evaluation requires a second method's evaluation outputs. PACT performs no gradient training; evaluation refolding dominates the cost.

Stage	Unit cost	Per antigen
Backbone generation (RFdiffusion, GPU)	14.5 s / backbone	256 backbones
Sequence decoding (AbMPNN, GPU)	<0.5 s / design	4,096 designs
Fitting-oracle refolding (IgFold, CPU)	2.1 s / design	2,048 designs ≈ 1.2 CPU-h
Fitting-oracle interface scoring (CPU)	9.4 s / design	2,048 designs ≈ 5.3 CPU-h
Coefficient estimation (Equation (9))	<1 s	closed form
Evaluation refolding (RF2, GPU)	30–49 s / design	2,048 designs ≈ 17–28 GPU-h
Interface scoring (Rosetta, CPU)	9.4 s / design	2,048 designs
Hardware (CPU)	AMD EPYC 7R13	
Hardware (GPU)	NVIDIA V100 / A100 / H100	

Coefficient fitting from existing decodes costs about 6.5 CPU-hours, whereas RF2 evaluation costs 17–28 GPU-hours per method and antigen. The efficiency claim concerns fitting; evaluation remains expensive. HER2 has the highest per-design cost because its modeled complex is the largest (675 residues).

D.8 Full-Parameter Training Objectives

We test six objectives on the same diagnostic data and reference policy: advantage regression (Equation (4)); DPO-style pairwise training using within-context reward orderings (Rafailov et al., 2023); advantage-weighted regression (AWR) (Peng et al., 2019); filtered SFT on the top reward quartile in each context (cf. Dong et al., 2023); position-weighted cross-entropy over positions that vary within a context; and a per-position KL-tilted target, $\mathrm{softmax}(\log \pi_{\mathrm{ref}}(\cdot \mid q) + \bar{A}(q,\cdot)/\beta)$. Across the runs collectively, learning rates range from 5×10^{-7} to 5×10^{-5}, gradient clipping ranges from 10 to effectively unclipped, β ranges from 0.1 to 3, and both soft-value and mean-reward baselines are tested. This is an uneven collection of configurations, rather than a full factorial sweep for every objective.

Offline regression targets. For the n candidates of a context, the two baselines used in advantage regression are

$$\hat{V}_{\text{soft}}(C) = \beta \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{r(C,\hat{S}_i)/\beta}\right), \qquad \hat{V}_{\text{mean}}(C) = \frac{1}{n}\sum_{i=1}^{n} r(C,\hat{S}_i).$$

Both are shared within a context. The first is a sample log-mean-exp, motivated by the soft value in Equation (2); the second centers the sampled rewards. The empirical regression averages the squared residual in Equation (4) over the available offline candidates. Candidates generated at low temperature and in random orders are not draws from the unit-temperature, fixed-order distribution used for checkpoint likelihood scoring. Thus, their sample log-mean-exp does not in general estimate that scoring distribution's soft value. The gradient identities in Appendix B apply to the fixed candidate set regardless of how it was generated.

Filtered-SFT fitting. For each antigen, filtered SFT imitates the top-quartile designs from the fitting split for 60 epochs. A 10% internal validation split leaves 116 of the 128 fitting backbones for training. No evaluation backbone appears in this input pool.

D.9 Checkpoint Probe Protocol

Training trajectories and sample. We analyze one 60-epoch mean-baseline advantage-regression run per antigen, initialized from the original AbMPNN reference. Each run uses 128 fitting-pool backbones, with a 10% internal validation split, 16 offline candidates per backbone, $\beta = 1$, AdamW with nominal learning rate 5×10^{-6}, weight decay 10^{-5}, batch size one, no gradient accumulation, gradient-norm clipping at 100, and seed 42. We use initialization and saved epochs 10, 20, 30, 40, 50, and 60. The optimizer states record 1,160 updates at epoch 10 and 6,960 at epoch 60; the stored learning rate is 5×10^{-8} at epoch 60.

Before inspecting the formal-study results, we select 12 evenly spaced backbone IDs from each half of the pool: 0, 12, 23, 35, 46, 58, 69, 81, 92, 104, 115, and 127 for held-out probes, and these IDs plus 128 for fitting-pool probes. The latter include the trainer's internal validation subset, whose exact membership is not saved, so we do not call every fitting-pool probe a training example. With all 16 candidates retained, the analysis covers 72 backbones, 504 checkpoint–backbone measurements, and 8,064 candidate gradients.

These trajectories are distinct from the six-objective sweep in Appendix E.1, which uses a different reference checkpoint and training-time correlation measurements. We also cannot establish the exact correspondence between these saved trajectories and the advantage-regression checkpoints used to generate the refolding comparisons in Table 1. The present results characterize the specified optimization runs; they are not a mechanistic attribution for those refolded designs.

Fixed scoring conditions. We reuse each backbone's 16 offline candidate sequences and the rewards supplied to its training run. All candidates are scored on the same canonical backbone coordinates, with identical framework and target sequences, design masks, and residue encodings. We fix a shared decoding order per backbone across candidates and checkpoints, disable coordinate noise, and score in evaluation mode. The probe sums the model's native conditional log-probabilities over designable CDR positions without generation-temperature scaling. It uses the 21-output forward scores (20 canonical residues plus the unknown-residue token `X`), with no additional omission mask or renormalization. These are scoring-model likelihoods, distinct from the low-temperature generation distribution ($\tau = 10^{-4}$ in Appendix D.4). The reference remains the original AbMPNN model throughout.

The mean baseline is computed in float64 from the complete reward vector, whereas historical training centered float32 rewards. These are controlled probes, not bitwise replays of historical losses. Candidate, coordinate, reward, source, decoding-order, and checkpoint hashes are recorded. For every checkpoint we save the complete 16×16 matrices K and PKP, their norms and spectra, relative log-ratios, residuals, and the two response terms in Equation (17).

Controlled optimizer steps. At epochs 0, 30, and 60, we take independent SGD and AdamW steps at a checkpoint-specific probe learning rate and half that rate, resetting weights and optimizer state

before each probe. Later probes use the checkpoint's stored learning rate; AdamW also restores the saved moments and step counter, without advancing the scheduler. SGD has no momentum or decay. Initialization uses fresh optimizer states and learning rate 5×10^{-7}, matching the original diagnostic rather than the historical warmup rate of 10^{-8}. Both optimizers use gradient clipping at 100. There are 864 controlled updates in total.

Each probe uses its own context's rewards. A held-out probe is therefore a counterfactual local update, not a test of generalization after updating on another context. Probe weights are discarded after measurement and never feed subsequent checkpoints or design generation. For the actual rounded parameter increment d_t, we compare the centered scaled log-ratio change with $\beta P J_t d_t$, recording correlation, relative error, and the bound in Equation (19). We separately compare its direction with the raw-kernel response $PK_t e_t$.

Numerical precision and validation. Gradients and parameter updates use float32, with kernel accumulation in float64. At epoch 60, small learning rates make likelihood differences sensitive to float32 subtraction. We also rescore the same endpoint parameter vectors using float64 neural-network arithmetic, retaining the original float32 geometric features and neighbor selection. Positional inputs are cast to the layer's arithmetic type; neither model parameters nor endpoint increments are changed. The first-order prediction still uses the float32 gradient and actual rounded increment. We report both readouts, separating numerical accuracy from changes with training stage.

Direct loss backpropagation agrees with the weighted sum of candidate gradients in the pilot to relative error at most 1.13×10^{-6}. Independent recomputation verifies the saved kernel spectra, residual decomposition, optimizer-state alignment, and first-order predictions; recorded hashes are checked for consistency across outputs. The formal analysis uses only CPUs: 72 tasks with four cores each, totaling 21.65 allocated CPU-hours.

E Full-Parameter Training Diagnostics

This section groups the six-objective sweep, the original initialization measurements, and the later checkpoint results. Their derivations are in Appendix B, and the checkpoint protocol is in Appendix D.9. The studies use different reference checkpoints and measurement protocols, so their summaries are kept separate.

E.1 Six-Objective Sweep and Refolded Comparisons

Objective definitions and training coverage are given in Appendix D.8.

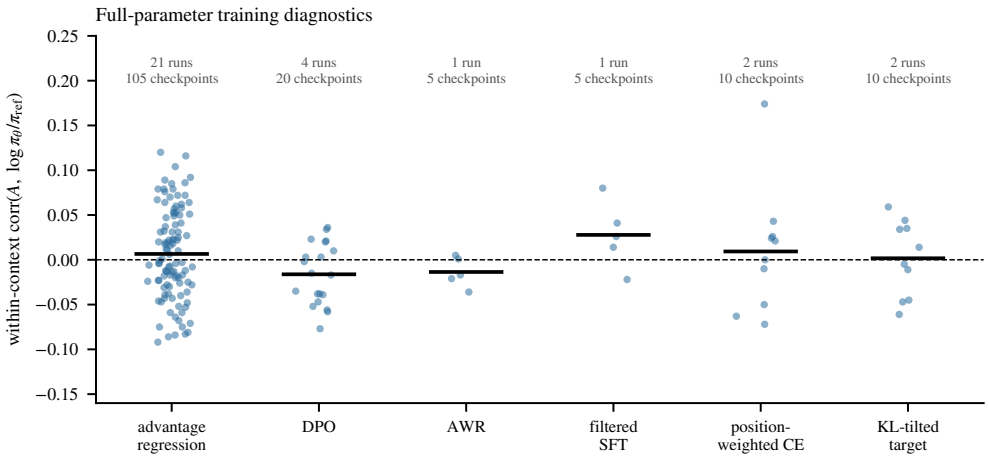


Figure 7: **Little within-context reward alignment under the tested objectives.** Each point is one recorded epoch: 155 measurements from 31 runs across six objectives. Black bars mark per-objective means. Epochs from the same run are dependent, so the plot is descriptive rather than a confidence-interval analysis. Coverage is uneven, with 21 advantage-regression runs and one each for AWR and filtered SFT. The filtered-SFT models evaluated by refolding were fitted separately on each antigen.

Within-context alignment. The correlation between centered reward and fitted log-ratio remains close to zero on average (Figure 7). The per-objective means are $+0.007$ for advantage regression (21 runs), -0.016 for DPO (4), -0.014 for AWR (1), $+0.028$ for filtered SFT (1), $+0.009$ for position-weighted cross-entropy (2), and $+0.002$ for the KL-tilted target (2). Across all 155 recorded epochs, the mean is 0.004. The fixed-order checkpoint probes in Appendix D.9 use a different reference and measurement protocol and are reported separately. These summaries describe the observed optimization behavior; repeated epochs are not independent experimental replicates.

Loss reduction and mean shifts. At initialization the policy log-ratio is zero, so the advantage-regression loss equals $\mathbb{E}[A^2] = 4.80$. The best recorded value is 4.13, a 14% reduction, followed by an increase in loss. This limited improvement does not produce a clear reward-aligned ranking. In a separate control, constant targets of $+1$ and -1 move the mean log-ratio to $+1.16$ and -0.88, respectively. The optimizer can therefore change average likelihood, even though it fails to recover the reward ordering in these runs. A nonzero constant log-ratio cannot hold over the entire normalized sequence distribution, since it would require $1 = e^c$. It can hold on a sampled subset, however, so the control tests movement on those sampled designs rather than attainment of a globally valid constant-ratio policy.

Held-out performance after filtered SFT. Despite closely fitting the training designs, the resulting models lose both reward and SC on every held-out target (Table 13). On HER2 and PD1, all three reward components decrease. On IL7R, the loss comes mainly from SC and hydrogen bonds; the buried-area difference is small (-6.3 Å^2, $p = 0.68$).

Direct paired comparisons favor PACT over filtered SFT by 1.831, 0.987, and 1.588 reward units on HER2, IL7R, and PD1, respectively. The corresponding SC differences are $+0.068$, $+0.066$, and $+0.086$, and PACT has higher reward on 87%, 73%, and 84% of backbones. All three direct reward tests pass Bonferroni correction ($p \leq 1.1 \times 10^{-15}$). Thus, the compact bias improves held-out design quality where near-exact full-parameter imitation degrades it.

Direct comparison with advantage regression. Advantage regression has higher held-out reward than filtered SFT on every target, making it the stronger full-parameter comparator. We therefore test PACT against it directly. Table 6 reports both direct contrasts: each method is averaged over its available designs within backbone, and the paired backbone differences are tested with a two-sided normal approximation. PACT exceeds advantage regression on HER2 and IL7R under a three-target correction ($\alpha' = 0.0167$). The PD1 difference is unresolved. The corresponding comparison against filtered SFT favors PACT on all three targets.

Table 6: Direct paired contrasts between PACT and each evaluated full-parameter policy, on 128 held-out backbones with up to 8 designs each. Positive values favor PACT. Intervals are normal 95% CIs over backbone means; p is two-sided. Each method column is its own three-target Bonferroni family ($\alpha' = 0.0167$); **bold** marks contrasts that pass. "bb. better" is the fraction of backbones on which PACT has the higher mean reward.

Comparison	Target	Δ reward	95% CI	p	bb. better
PACT − advantage regression	HER2	$\mathbf{+0.491}$	$[+0.267, +0.715]$	1.8×10^{-5}	64%
	IL7R	$\mathbf{+0.303}$	$[+0.095, +0.512]$	4.4×10^{-3}	56%
	PD1	$+0.072$	$[-0.131, +0.275]$	0.49	49%
PACT − filtered SFT	HER2	$\mathbf{+1.831}$	$[+1.520, +2.141]$	5.8×10^{-31}	87%
	IL7R	$\mathbf{+0.987}$	$[+0.746, +1.228]$	1.0×10^{-15}	73%
	PD1	$\mathbf{+1.588}$	$[+1.329, +1.847]$	3.3×10^{-33}	84%

The corresponding SC contrasts against advantage regression are $+0.0138$, $+0.0164$, and $+0.0040$ ($p = 0.011$, 0.012, and 0.52). The HER2 and IL7R values are below 0.0167; PD1 is unresolved. We treat these secondary SC contrasts as exploratory. On HER2, advantage regression contributes 881 of 1,024 designs (Table 13); the contrast averages each method's available designs within backbone before pairing, as elsewhere in this comparison.

The filtered-SFT evaluation uses final imitation checkpoints. It does not determine whether early stopping or stronger regularization would prevent degradation. Advantage regression also has refolded evaluations (Table 13); the other four objectives are assessed only through the likelihood-ratio diagnostic.

E.2 Original initialization measurements

We measure the reference policy on 24 held-out backbones per antigen, with 16 candidates per backbone. These measurements characterize initialization, not a training trajectory. The diagnostic backbones are separate from those used to fit the composition coefficients. We compute $g_i = \nabla_\theta \log \pi_{\mathrm{ref}}(\hat{S}_i \mid C)$ over all network parameters, with the log-likelihood restricted to designable CDR positions. All candidates are scored with the same fixed decoding order, and dropout is disabled by placing the model in evaluation mode. These controls ensure that gradient differences reflect the candidates rather than randomness in the forward pass. Table 7 reports medians over backbones. The leading-eigenvector alignment is $|v_1^\top \mathbf{1}|/\sqrt{n}$, with $\|v_1\| = 1$.

Table 7: Gradient-kernel geometry on same-backbone candidates ($n = 16$ candidates, 24 backbones per antigen, medians). $\|PKP\|_2/\|K\|_2$ is the fraction of the kernel norm that survives centering. The absolute centered norm bounds the initial response in Equation (18); this ratio compares that bound with the one obtained from the full kernel. Sequence lengths and distances count amino-acid residues at the annotated designable positions.

	HER2	**IL7R**	**PD1**
$\|PKP\|_2/\|K\|_2$	0.183	0.138	0.193
$\lambda_1/\mathrm{tr}\, K$	0.664	0.728	0.645
alignment of v_1 with $\mathbf{1}$	0.993	0.995	0.993
mean pairwise gradient cosine	0.630	0.698	0.605
designable positions	31	31	31
mean pairwise Hamming distance	4.1	4.3	4.1
positions varying across the 16 candidates	10.5	10.5	10

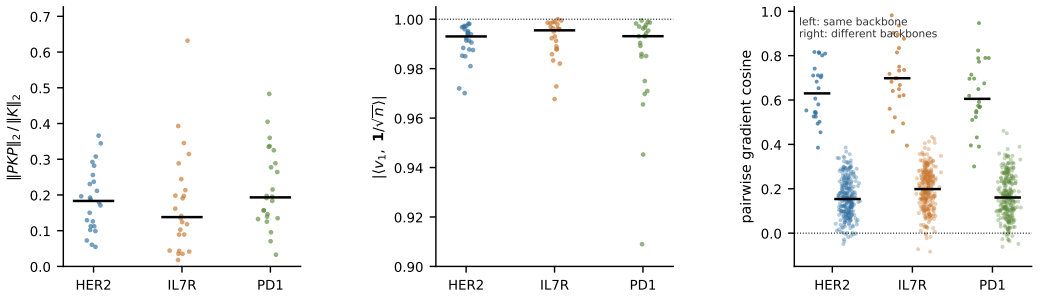
The leading eigenvector is nearly constant, and its eigenvalue accounts for 64–73% of the kernel trace. Centering retains 14–19% of the spectral norm, so the initial response to centered rewards is bounded more tightly than the full kernel norm suggests. The nonzero centered norm still permits first-order changes in relative likelihoods.

Gradient alignment also weakens as candidate sequences diverge. Across 72 backbones, the mean pairwise gradient cosine decreases with mean pairwise residue Hamming distance ($r = -0.69$). This is a descriptive association across the sampled backbones; it complements the same-backbone and different-backbone comparison in Figure 8c. The cross-backbone pairs in that figure share backbones, so their cosine similarities are descriptive observations rather than independent replicates.

E.3 Validating the initial response with real rewards

The checkpoint study also measures initialization using the actual candidate rewards. Table 8 compares the direction PK_0PA with the centered log-ratio change under independent SGD and AdamW probes. Both optimizers use fresh state, fixed decoding order, evaluation mode, and learning rate 5×10^{-7}. The sample includes 12 held-out and 12 fitting-pool contexts per antigen, as specified in Appendix D.9.

SGD closely follows the kernel direction. AdamW also has high directional agreement at initialization, while its preconditioning changes the increment. The later-checkpoint analysis in Appendix E.5 therefore tests predictions using the actual parameter increment and compares them separately with the raw-kernel response.



(a) Retained kernel norm (b) Leading direction (c) Gradient similarity

Figure 8: **Gradient geometry at initialization.** We evaluate 16 candidates on each of 24 backbones per antigen with fixed decoding order. (a) Fraction of kernel spectral norm retained after removing the candidate-shared component. (b) Alignment of the leading eigenvector with the constant direction $\mathbf{1}$. Points in (a,b) denote backbones. (c) Gradient cosine similarity within the same backbone (dark) and across backbones (light). Black bars show medians. Gradients align more strongly within backbones, while centering substantially reduces the kernel norm.

Table 8: **Initial kernel direction against actual optimizer steps.** Median correlation between PK_0PA and the realized centered log-ratio change, over 24 contexts per antigen. Targets are the observed centered rewards. These are controlled single-context probes; AdamW need not follow the SGD step magnitude.

Optimizer	HER2	IL7R	PD1
SGD	0.9999	0.9999	0.9998
AdamW	0.9966	0.9971	0.9960

E.4 Checkpoint geometry and reward fitting

Persistent structure with measurable drift. Figure 3 and Table 9 show that the small centered/full norm ratio persists through the saved training stages. The centered kernel is not fixed: its final drift relative to initialization is 14–21% across antigen/split medians. Absolute centered norms remain substantial, so a small ratio should not be interpreted as vanishing gradients. The final mean-coupling response exceeds the centered-residual response in 52 of 72 contexts. Thus, the shared gradient component remains prominent while its coupling to relative updates becomes important after initialization.

Reading Figure 3. Panel (a) expresses the retained spectral norm as a percentage, $100\|PK_tP\|_2/\|K_t\|_2$. In panel (b), the numerator and denominator denote the norms of $\eta_t\bar{\ell}_tPK_t\mathbf{1}$ and $\eta_tPK_tPe_t$, respectively, from Equation (5). Their ratio exceeds one when the mean-shift term is larger; it compares the two local response terms, not their alignment with reward or their contribution to a full AdamW step. Panel (c) first centers log-ratios and rewards within each backbone, then pools candidates within each antigen/split. It reports the resulting reward-fit MSE divided by its initial value, so values below one indicate improvement. Panels (a,b) show backbone medians.

Table 9: **Kernel geometry at epoch 60.** Entries are medians over 12 backbones per antigen/split. $H_t = PK_tP$. Absolute norms are in units of 10^4; drift is $\|H_{60} - H_0\|_2/\|H_0\|_2$. Coupling/relative is the ratio of the response-term norms in Equation (5); the common step-size factor cancels.

Target	Split	$\|H_0\|_2$	$\|H_{60}\|_2$	Centered/full	Drift	Coupling/relative
HER2	Held-out	8.096	7.534	0.161	0.165	1.329
HER2	Fitting pool	7.655	6.498	0.133	0.163	1.916
IL7R	Held-out	5.371	4.439	0.125	0.209	2.027
IL7R	Fitting pool	5.854	5.284	0.121	0.160	2.125
PD1	Held-out	6.639	6.022	0.177	0.174	1.505
PD1	Fitting pool	7.470	7.273	0.164	0.143	1.934

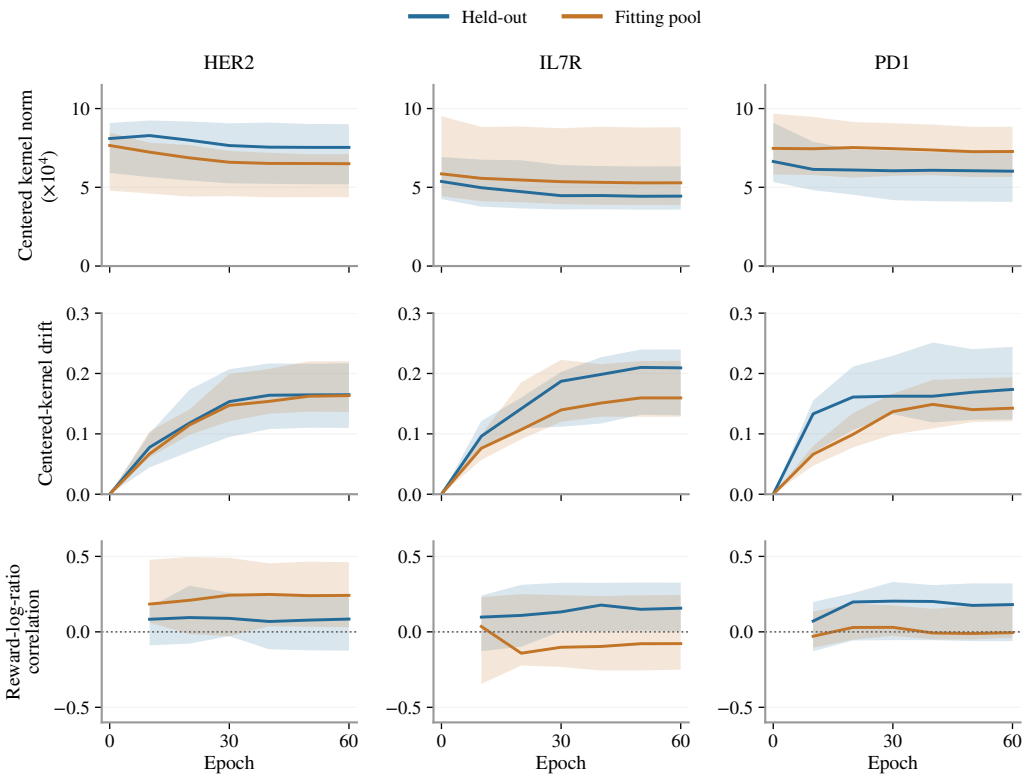


Figure 9: **Absolute kernel scale, drift, and reward alignment across training.** Lines show medians and bands the interquartile range over 12 backbones per antigen/split; they are not confidence intervals. Drift is measured relative to the initial centered kernel. Reward–log-ratio correlation is undefined at initialization because all log-ratios are zero. The fitting pool includes the original internal validation subset.

Relative changes do not consistently fit reward differences. The final log-ratios are not uniformly uncorrelated with reward (Figure 9 and Table 10). Some groups show positive alignment, and the pooled centered log-ratio norm is 31–46% of the advantage norm. Nevertheless, the final centered error exceeds its initial value in five of six groups, and the remaining group reduces centered MSE by only 1.05%. These probes support weak and uneven reward fitting, rather than the absence of any reward-responsive direction. For example, a controlled initial AdamW step has median reward–update correlation 0.59–0.67 across the six groups, so a locally aligned response does exist.

Table 10: **Reward fitting at epoch 60.** Median r is the backbone-level median reward–log-ratio correlation. Pooled statistics first center each context's vectors, then concatenate them within antigen/split. The residual ratio is $\|P(A-\ell)\|/\|PA\|$ on these concatenated vectors. MSE reduction is relative to initialization; negative values indicate a larger final error. All entries are descriptive.

Target	Split	Median r	Pooled r	Residual ratio	MSE reduction
HER2	Held-out	0.085	0.047	1.052	-10.76%
HER2	Fitting pool	0.242	0.223	1.003	-0.52%
IL7R	Held-out	0.157	0.132	1.025	-5.06%
IL7R	Fitting pool	-0.078	-0.004	1.102	-21.40%
PD1	Held-out	0.181	0.170	0.995	+1.05%
PD1	Fitting pool	-0.005	0.021	1.071	-14.79%

E.5 LOCAL PREDICTIONS AT LATER CHECKPOINTS

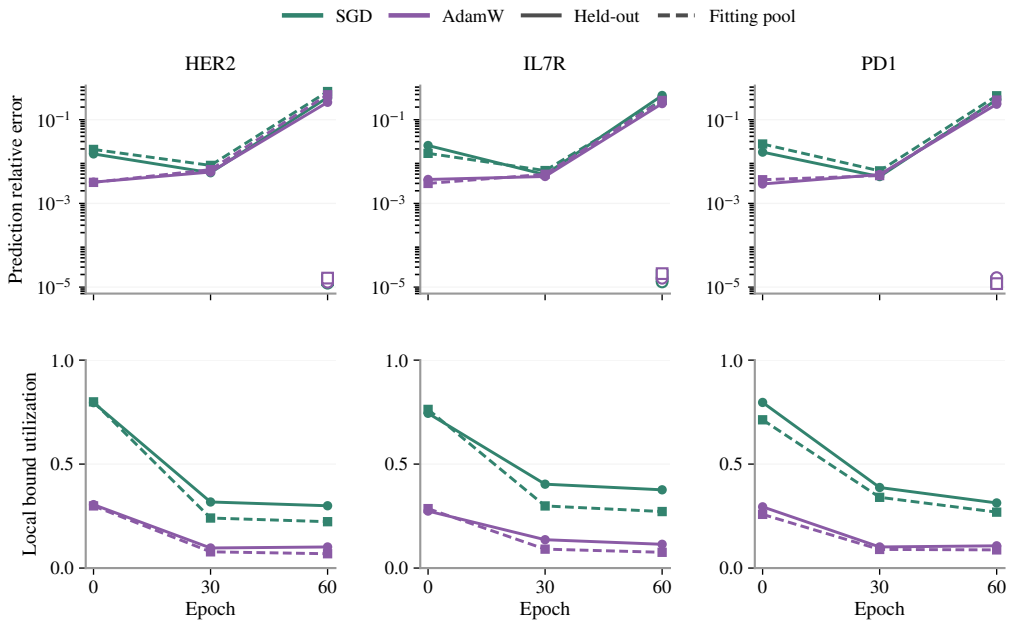


Figure 10: **Local prediction accuracy and bound utilization.** Curves show medians over 12 backbones per antigen/split at nominal probe learning rate. Top: connected points use float32 likelihood readout throughout; isolated open points at epoch 60 use float64 readout of the same parameter endpoints. Bottom: first-order prediction norm divided by the optimizer-independent bound in Equation (19). Later AdamW probes restore saved moments and learning rates; initial probes use fresh states at 5×10^{-7}. Each probe uses its context's own rewards, so held-out probes are not post-update generalization tests.

Direction and magnitude. At epoch 30, the first-order prediction from the actual AdamW increment has median relative error 0.44–0.64% across the six antigen/split groups at nominal probe learning rate. At epoch 60, float64 readout reduces the apparent discrepancy to median relative errors of 1.2×10^{-5}–2.1×10^{-5} for both optimizers at the stored learning rate (Table 11 and Figure 10). This supports the local expansion beyond initialization. The large float32 discrepancies at epoch 60 arise from numerical readout precision and should not be read as failure of the local prediction.

Table 11: **Epoch-60 local update checks.** Backbone medians, with float64 likelihood readout and float32 gradients and parameter increments. Error is $\|P\Delta\ell - \beta PJd\|/\|P\Delta\ell\|$. SGD and AdamW use the stored learning rate; AdamW $\times\frac{1}{2}$ halves it. The last column correlates the raw-kernel response PKe with the realized centered log-ratio change under AdamW at the stored rate, distinguishing kernel direction from the prediction that uses the actual parameter increment.

		Prediction relative error			Raw-kernel
Target	Split	SGD	AdamW	AdamW $\times\frac{1}{2}$	correlation
HER2	Held-out	1.24×10^{-5}	1.31×10^{-5}	7.62×10^{-6}	0.978
HER2	Fitting pool	1.50×10^{-5}	1.64×10^{-5}	1.34×10^{-5}	0.977
IL7R	Held-out	1.33×10^{-5}	1.63×10^{-5}	8.18×10^{-6}	0.986
IL7R	Fitting pool	1.88×10^{-5}	2.10×10^{-5}	1.37×10^{-5}	0.955
PD1	Held-out	1.56×10^{-5}	1.69×10^{-5}	1.15×10^{-5}	0.988
PD1	Fitting pool	1.33×10^{-5}	1.21×10^{-5}	8.53×10^{-6}	0.975

Optimizer dependence and scope. The raw-kernel response also tracks AdamW reasonably well on average: its median correlation across all 72 contexts is 0.974 at epoch 30 and 0.979 at epoch 60,

but individual values fall to 0.592 and 0.671, respectively, at nominal probe learning rate. Accurate prediction using the actual increment therefore does not imply that the current kernel alone specifies AdamW's direction or step magnitude. Figure 10 also reports utilization of the local norm bound, defined as $\|\beta PJd\|/(\beta\sqrt{\|PKP\|_2}\|d\|)$.

The evidence covers one training run per antigen and sparse checkpoints, with deterministic candidate scoring and independent single-context probes. It does not recover historical minibatches, stochastic draws, or the full parameter path, nor establish generality across training seeds or all six objectives. Together with the persistent geometry, it supports an account of later local responses; it does not certify the path bound or prove that the observed geometry causes long-run optimization failure.

F Additional Experimental Results

F.1 Detailed Primary and Method Comparisons

Table 12 gives the numerical results behind Figure 4. Table 13 retains the absolute reference rewards and reward p-values for the compact method comparison in Table 1.

Table 12: **Main results.** PACT versus the AbMPNN reference policy on RF2-refolded structures, 128 held-out backbones $\times$ 16 decodes per antigen. Designs from one backbone are not independent, so every Δ is the mean of within-backbone paired differences and all inference is over the $n = 128$ *backbones*. **Bold** marks the better-performing method and its Δ, and only where the difference is significant at $\alpha = 0.05$. Reward and SC also meet their respective three-test Bonferroni thresholds; the other components are descriptive secondary analyses. Fitting and evaluation backbones are disjoint.

Antigen	Metric	AbMPNN	PACT	Δ	95% CI	p	bb. improved
HER2	Reward	12.429	**13.063**	**+0.634**	$[+0.456, +0.811]$	2.6×10^{-12}	71%
	SC	0.3970	**0.4128**	**+0.0157**	$[+0.0076, +0.0239]$	1.6×10^{-4}	63%
	ΔSASA	1734.8	**1836.6**	**+101.9**	$[+71.6, +132.1]$	4.1×10^{-11}	73%
	Hbonds	2.361	**2.577**	**+0.215**	$[+0.085, +0.346]$	1.2×10^{-3}	66%
IL7R	Reward	9.937	**10.335**	**+0.399**	$[+0.244, +0.553]$	4.3×10^{-7}	66%
	SC	0.4070	**0.4296**	**+0.0226**	$[+0.0128, +0.0323]$	6.2×10^{-6}	63%
	ΔSASA	1109.4	**1135.8**	**+26.5**	$[+5.5, +47.4]$	1.3×10^{-2}	55%
	Hbonds	**1.586**	1.445	**-0.141**	$[-0.238, -0.044]$	4.4×10^{-3}	39%
PD1	Reward	9.859	**10.175**	**+0.315**	$[+0.131, +0.500]$	8.2×10^{-4}	59%
	SC	0.3808	**0.3973**	**+0.0165**	$[+0.0060, +0.0270]$	2.1×10^{-3}	65%
	ΔSASA	1172.5	1196.8	$+24.3$	$[-1.7, +50.3]$	6.7×10^{-2}	59%
	Hbonds	1.597	1.613	$+0.017$	$[-0.097, +0.131]$	0.78	45%

Comparisons with other generators. PACT has the higher reward point estimate in all twelve comparisons. It significantly exceeds AbLang and nanoBERT on all three targets and AntiFold on two, while the differences from ESM-IF do not pass correction ($p = 0.0084$, 0.41, and 0.023). Non-significant differences do not establish equivalence. The sequence-only results should be interpreted in light of their limited conditioning.

Table 13: **Full comparison of sequence-generation methods.** Paired changes from AbMPNN on 128 backbones with up to 8 designs each, with the absolute reference reward in each target header. The two-sided p-values compare each method's reward with AbMPNN, using paired t-tests for advantage regression and a normal approximation otherwise. ΔE is the change in unrelaxed Rosetta separated energy (REU; lower is favorable), with exploratory normal 95% confidence intervals. These are not direct tests between methods. Each method is averaged within backbone over its available designs. HER2 advantage regression has 881/1,024 refolds (86%), spanning all backbones; its reference retains all eight designs. AbLang and nanoBERT receive no antigen context.

Method	Δ reward	p (reward)	ΔSC	ΔE	95% CI (ΔE)
HER2 (AbMPNN reward: 12.455)					
PACT	$+0.574$	5.7×10^{-8}	$+0.0149$	-424	$[-736, -113]$
Advantage regression	$+0.084$	0.42	$+0.0010$	-100	$[-402, +201]$
Filtered SFT	-1.256	7.7×10^{-18}	-0.0527	$+305$	$[-214, +824]$
ESM-IF	$+0.121$	0.46	-0.0136	$+1001$	$[+282, +1719]$
AntiFold	-0.513	3.7×10^{-3}	-0.0171	$+696$	$[+66, +1327]$
AbLang	-0.318	0.068	-0.0096	$+379$	$[-273, +1032]$
nanoBERT	$+0.058$	0.75	$+0.0084$	-113	$[-613, +386]$
IL7R (AbMPNN reward: 9.901)					
PACT	$+0.435$	3.5×10^{-6}	$+0.0265$	-430	$[-697, -163]$
Advantage regression	$+0.132$	0.056	$+0.0101$	-122	$[-360, +117]$
Filtered SFT	-0.552	4.7×10^{-6}	-0.0392	$+2192$	$[+1542, +2842]$
ESM-IF	$+0.324$	7.2×10^{-3}	$+0.0301$	-701	$[-1089, -313]$
AntiFold	-0.057	0.67	$+0.0120$	-404	$[-858, +51]$
AbLang	-0.180	0.16	-0.0101	$+556$	$[-57, +1168]$
nanoBERT	-0.027	0.84	-0.0040	-14	$[-539, +510]$
PD1 (AbMPNN reward: 9.868)					
PACT	$+0.231$	0.037	$+0.0125$	-279	$[-612, +54]$
Advantage regression	$+0.159$	0.068	$+0.0085$	-100	$[-380, +179]$
Filtered SFT	-1.357	2.9×10^{-24}	-0.0730	$+1386$	$[+858, +1914]$
ESM-IF	-0.030	0.82	-0.0003	-80	$[-515, +356]$
AntiFold	$+0.123$	0.40	$+0.0184$	-343	$[-750, +64]$
AbLang	-0.360	0.035	-0.0050	$+25$	$[-537, +588]$
nanoBERT	-0.483	2.1×10^{-3}	-0.0237	$+653$	$[+85, +1221]$

F.1.1 Metrics Outside the Reward

The primary comparison evaluates a fitted interface score. To characterize changes outside that score, we analyze archived separated energies and structural predictions. These analyses reuse existing outputs and are exploratory. For each metric we form a mean difference within each backbone, then report the mean and its normal 95% confidence interval across backbones, using the same normal approximation as the primary analysis. The direction of improvement is stated separately for each metric.

Separated energy in the main 16-design comparison. We extract `dG_separated` from the same RF2 score CSVs as the primary reward comparison and join PACT and reference designs by backbone and decode slot. There are 2,048 pairs over 128 backbones per target. As a check on cohort identity, this join reproduces the reference and PACT means, paired differences, standard errors, and confidence intervals of all four primary metrics in Table 12. The values are from unrelaxed poses, with no FastRelax pass. Raw energies are dominated by large positive values; the reductions in Table 14 should therefore be read as favorable changes in this Rosetta calculation, not as physical binding free energies. The separate relaxed medians in Table 21 are a different diagnostic.

Structural self-consistency on archived RF2 outputs. For each design we match the heavy-chain residue IDs and amino-acid sequence between its input structure and RF2 prediction. We then align the complete heavy-chain C_α traces using Kabsch superposition and compute RMSD. We reject mismatches instead of truncating chains to the shorter length. Reference and PACT values are

paired by backbone and decode slot before aggregation. This measures internal antibody-backbone agreement, not the antibody's pose relative to the antigen. Sample coverage is given with the results below. In contrast, the archived control summaries in Table 21 use alphabetical subsets and are not substitutes for this paired analysis.

Table 14: **Paired diagnostics for the main PACT–AbMPNN comparison.** Means give equal weight to each backbone and use only the common valid designs in each comparison. Lower is favorable for both metrics. Differences are PACT minus AbMPNN; intervals are unadjusted normal 95% CIs. Energy is in REU and scRMSD in Å. Coverage is stated below.

Target	Metric	AbMPNN	PACT	Δ	95% CI
HER2	ΔG_{sep}	3076.2	2738.2	−338.0	$[-584.6, -91.4]$
	scRMSD	1.654	1.620	−0.034	$[-0.085, +0.018]$
IL7R	ΔG_{sep}	1990.4	1596.4	−394.0	$[-641.9, -146.1]$
	scRMSD	1.575	1.540	−0.034	$[-0.092, +0.023]$
PD1	ΔG_{sep}	2722.5	2296.9	−425.6	$[-714.7, -136.6]$
	scRMSD	1.683	1.641	−0.042	$[-0.098, +0.015]$

The energy analysis uses 2,048 paired designs over 128 backbones per target. The HER2 scRMSD comparison uses 1,978 paired designs over 128 backbones. The IL7R scRMSD comparison uses 2,048 paired designs over 128 backbones. The PD1 scRMSD comparison uses 2,048 paired designs over 128 backbones. Only structures available in the archived RF2 output directories and passing the residue-ID and sequence checks enter scRMSD. The energy CSVs retain scores for the complete planned set even where a PDB is unavailable.

Structural comparison at the eight-design budget. We apply the same scRMSD definition to every method in Table 13, using decode slots 0–7 from the original method-comparison score files. Each input/RF2 pair must pass the residue-ID and sequence checks above. All methods span 128 backbones per target. HER2 PACT has 971 available structure pairs out of 1,024 scored designs; HER2 advantage regression has 881 scored designs and structure pairs. All other comparisons have 1,024. As in the other columns of Table 1, we average each method's available designs within backbone before pairing backbone means. This differs from the common-slot pairing in the main 16-design comparison. The numerical export also includes common-slot sensitivity analyses and direct paired comparisons of PACT with each alternative method. Table 15 reports the method means and direct PACT comparisons.

Table 15: **Structural self-consistency across sequence-generation methods.** Heavy-chain scRMSD in Å at the eight-design budget, with equal weight per backbone. Bracketed values are exploratory, unadjusted normal 95% CIs from paired backbone means. The third column compares each method with AbMPNN; negative values in the last column favor PACT over the named method. Means use available structures, with coverage specified in the text.

Method	Mean	Change from AbMPNN	PACT minus method
HER2 (AbMPNN: 1.648 Å)			
PACT	1.625	−0.022 [−0.079, +0.035]	—
Advantage regression	1.639	−0.008 [−0.041, +0.025]	−0.014 [−0.068, +0.040]
Filtered SFT	2.020	+0.372 [+0.291, +0.453]	−0.395 [−0.478, −0.311]
ESM-IF	1.536	−0.112 [−0.199, −0.024]	+0.090 [−0.003, +0.182]
AntiFold	1.837	+0.189 [+0.099, +0.280]	−0.212 [−0.310, −0.114]
AbLang	2.451	+0.803 [+0.680, +0.927]	−0.825 [−0.950, −0.701]
nanoBERT	2.157	+0.510 [+0.419, +0.601]	−0.532 [−0.625, −0.439]
IL7R (AbMPNN: 1.575 Å)			
PACT	1.534	−0.041 [−0.099, +0.017]	—
Advantage regression	1.552	−0.023 [−0.060, +0.014]	−0.018 [−0.072, +0.036]
Filtered SFT	2.123	+0.548 [+0.472, +0.624]	−0.589 [−0.665, −0.514]
ESM-IF	1.565	−0.010 [−0.117, +0.097]	−0.031 [−0.127, +0.065]
AntiFold	1.731	+0.156 [+0.062, +0.250]	−0.197 [−0.297, −0.098]
AbLang	2.568	+0.993 [+0.865, +1.121]	−1.034 [−1.173, −0.895]
nanoBERT	2.265	+0.690 [+0.589, +0.790]	−0.731 [−0.830, −0.631]
PD1 (AbMPNN: 1.694 Å)			
PACT	1.648	−0.046 [−0.107, +0.016]	—
Advantage regression	1.664	−0.030 [−0.062, +0.003]	−0.016 [−0.070, +0.038]
Filtered SFT	2.207	+0.512 [+0.423, +0.601]	−0.558 [−0.642, −0.474]
ESM-IF	1.641	−0.053 [−0.168, +0.062]	+0.007 [−0.100, +0.114]
AntiFold	1.915	+0.221 [+0.134, +0.309]	−0.267 [−0.353, −0.180]
AbLang	2.549	+0.855 [+0.707, +1.004]	−0.901 [−1.050, −0.751]
nanoBERT	2.356	+0.661 [+0.550, +0.772]	−0.707 [−0.813, −0.601]

F.2 Fitted Composition Coefficients

Table 16 reports the per-residue coefficients $\hat{b}_a$ of Equation (9), fitted independently for each antigen on its 128 fitting backbones (disjoint from the evaluated backbones). The three targets agree only partially. Across the 17 residues estimated on all three targets, the coefficient correlations are $r = +0.20$ (HER2/IL7R), $+0.27$ (HER2/PD1), and $+0.08$ (IL7R/PD1), and 7 of the 17 residues estimated on all three change sign between them. The shared pattern is limited: arginine is favored on all three and the small residues Gly, Ala, Ser and Thr are penalized on all three, indicating that these composition changes are associated with higher fitting rewards. The remaining preferences vary by target; the coefficients do not identify causal residue effects.

Table 16: Fitted per-residue coefficients $\hat{b}_a$ (top five favorable and unfavorable per antigen), estimated on fitting backbones disjoint from evaluation via Equation (9) using IgFold-refolded rewards.

Antigen	Favorable	Unfavorable
HER2	W +0.197, R +0.107, Q +0.074, K +0.062, E +0.048	V −0.085, S −0.034, T −0.021, A −0.018, G −0.012
IL7R	M +0.082, K +0.058, P +0.053, R +0.049, V +0.035	I −0.072, S −0.022, A −0.021, G −0.013, T −0.002
PD1	M +0.077, E +0.060, Y +0.048, R +0.042, I +0.040	P −0.069, Q −0.040, G −0.036, T −0.024, K −0.023

Interpretation of the estimator. Equation (9) predicts centered sequence reward from the identity of a single residue occurrence. It is not a joint least-squares fit of an additive sequence-reward model:

Table 17: Complete fitted coefficient vectors $\hat{b}$ (all 20 residues, three antigens). "—" marks a residue observed fewer than 150 times in the fitting set, which is assigned $\hat{b}_a = 0$; cysteine and histidine are essentially absent from designed CDRs, and tryptophan clears the threshold only on HER2. A printed 0.0000 is distinct from a dash: IL7R glutamine clears the threshold with 577 occurrences and its mean centered reward rounds to zero at this precision. The numeric vectors are given here in full; the machine-readable files and the analysis scripts that consume them will accompany the camera-ready code.

	HER2	IL7R	PD1		HER2	IL7R	PD1
A	−0.0183	−0.0211	−0.0127	M	+0.0134	+0.0815	+0.0772
C	—	—	—	N	+0.0202	+0.0064	+0.0258
D	+0.0179	+0.0204	+0.0078	P	+0.0050	+0.0532	−0.0688
E	+0.0484	+0.0060	+0.0598	Q	+0.0736	+0.0000	−0.0399
F	−0.0050	+0.0253	−0.0017	R	+0.1069	+0.0490	+0.0422
G	−0.0122	−0.0130	−0.0357	S	−0.0341	−0.0217	−0.0037
H	—	—	—	T	−0.0206	−0.0017	−0.0236
I	+0.0205	−0.0715	+0.0397	V	−0.0848	+0.0347	−0.0147
K	+0.0618	+0.0580	−0.0233	W	+0.1970	—	—
L	+0.0100	+0.0225	−0.0064	Y	+0.0127	+0.0328	+0.0477

it assigns the whole centered reward to each occurrence without controlling for the other residues present. Appendix C.3 derives the estimator and shows how co-occurrence affects it even under an exactly additive reward model. We assess its usefulness through the controls of Section 5 and the stability check below.

Stability of the fit. Residue occurrences are dependent because each sequence reward is reused across its variable positions, so their count does not measure the effective sample size. A separate diagnostic fits coefficients on alternating halves of all 256 backbone contexts, using a 60-occurrence threshold. The two fits correlate at 0.581 on HER2 and 0.620 on PD1 with IgFold, versus 0.240 on PD1 with RF2. This diagnostic includes evaluation contexts; the deployed vectors use only the 128 fitting backbones and the 150-occurrence threshold. It supports the choice of fitting oracle, while measuring stability over the full backbone pool.

F.3 Bias-Strength Sensitivity in an Earlier Pilot

Table 18 records an earlier HER2 pilot with 512 designs per method and a separately fitted composition vector. It illustrates why aggregate reward alone can be a misleading selection criterion. The highest SC is 0.417 at scale 10, close to 0.415 at scale 20. At scale 60, the reward gain is larger, but SC falls to 0.317 and 18.9% of designs bury more than 3,000 Å^2 of surface, compared with 2.1% for AbMPNN. Stronger steering can therefore increase the score by enlarging the interface while worsening packing.

Table 18: Archived HER2 bias-strength pilot (512 designs per method, RF2-refolded). Scales are the absolute multipliers recorded for the pilot vector; this vector and design set are separate from the main comparisons. Dashes mark unavailable summaries.

Pilot scale	Δ reward	SC	Designs > 3,000 Å^2	Interpretation
10	+0.665	0.417	—	usable
20	+0.760	0.415	1.8%	usable
30	+0.622	—	—	degrading
40	+0.115	0.364	1.6%	degraded
60	+1.313	0.317	18.9%	artifact (area inflation)
AbMPNN reference	—	0.401	2.1%	—

These archived pilot scales apply to that pilot's coefficient vector. They are not multiples of the scale 19 used with the saved IgFold refits, and the available record does not establish a corresponding sweep for those refits (Appendix D.4).

F.4 Richer Restricted Families

We also fit separate coefficient vectors for H1, H2, and H3, giving 60 coefficients and 57 effective degrees of freedom. This extension can express loop-specific preferences: leucine is favored in H1 ($+0.050$) but disfavored in H3 (-0.042), while tryptophan's largest coefficient is in H3 ($+0.100$).

In a separate paired HER2 comparison, the per-loop variant exceeds the shared bias by $+0.026$ reward units, but the difference is unresolved ($p = 0.41$). We retain the simpler shared bias because this experiment provides no clear benefit from the extension. The result does not establish equivalence or rule out gains from richer families on other targets.

F.5 Choosing the Steering Vector

Table 19 is the aggregate-reward table for the steering controls discussed in Section 5; Table 21 gives the archived IL7R diagnostic panel, whose secondary metrics were computed on smaller subsets than the interface scores.

Table 19: **Reward changes under different steering vectors.** All vectors are applied to frozen AbMPNN at $1/\beta = 19$. Changes from AbMPNN are averaged within backbone and tested across 128 held-out backbones, with eight designs each. **Bold** marks the largest gain among the three fitted-vector rows. Shuffling permutes the target's coefficients. The volume prior uses side-chain volume but retains the target's fitted support and L_1 norm. Transferred vectors use only source-target coefficients and retain their source norms. These are the comparisons summarized in Table 2; norm-matched variants are in Table 22.

Steering vector	**HER2**		**IL7R**		**PD1**	
	Δ reward	p	Δ reward	p	Δ reward	p
Shuffled (magnitude-matched)	-0.760	7.0×10^{-9}	-0.097	0.23	-0.174	0.069
Size prior (volume direction)	$+0.754$	1.8×10^{-9}	$+0.424$	4.0×10^{-6}	$+0.001$	1.00
The antigen's own $\hat{b}$	$+\mathbf{0.786}$	9.7×10^{-9}	$+0.306$	2.1×10^{-4}	$+0.287$	6.7×10^{-3}
Pooled from other antigens	$+0.410$	1.0×10^{-4}	$+\mathbf{0.639}$	1.5×10^{-13}	$+\mathbf{0.328}$	3.9×10^{-3}
One other antigen's	$+0.428$	9.7×10^{-5}	$+0.494$	1.3×10^{-8}	$+0.214$	0.075

Relation to logit sharpening. At $\tau = 10^{-4}$, decoding is effectively an argmax of $z_a + b_a/\beta$. Multiplying the reference logits by a positive scalar leaves their argmax unchanged for a fixed prefix and decoding order, whereas an additive residue bias can change the selected residue. This distinction in the greedy limit does not replace controls matched on the induced output distribution at finite temperature.

Direct contrasts between steering vectors. Table 20 compares the target-fitted vector directly with each control, using the same backbones and decoding budget. It outperforms the shuffled vector on all three targets. Against the volume prior, it improves PD1, while HER2 and IL7R remain unresolved. The pooled vector outperforms the target-specific vector on IL7R and underperforms it on HER2; the PD1 difference is unresolved. Together, these results show that useful residue preferences can transfer across targets, but do not isolate their benefit from a size prior on every antigen. The contrasts are exploratory and use one shuffle permutation.

F.6 Archived Steering Diagnostics

Table 21 reports the archived IL7R summaries. The interface columns use the eight-design bias-vector comparison; scRMSD and Sim use the subsets specified below. Compared with using the fitted vector, shuffling its coefficients gives lower SC, higher scRMSD, and less favorable relaxed ΔG_{sep}. These descriptive comparisons do not establish a matched improvement over AbMPNN: the structural and relaxed-energy columns use the smaller, separately sampled subsets below.

Table 20: Direct paired contrasts between the target-fitted vector $\hat{b}$ and each control, at the common multiplier $1/\beta = 19$ on 128 held-out backbones $\times$ 8 designs. Positive values favor the target-fitted vector. Two-sided normal tests on paired backbone means are exploratory and uncorrected.

	HER2		IL7R		PD1	
$\hat{b}$ minus	Δ reward	p	Δ reward	p	Δ reward	p
Shuffled	$+1.546$	3.8×10^{-29}	$+0.403$	2.8×10^{-5}	$+0.461$	4.1×10^{-5}
Size prior (volume)	$+0.032$	0.77	-0.118	0.24	$+0.286$	5.6×10^{-3}
Pooled	$+0.376$	2.3×10^{-3}	-0.333	2.6×10^{-5}	-0.041	0.71

The table distinguishes its two reference sets. Reward components are available for the 8-slot paired subset, but its relaxed ΔG_{sep}, scRMSD, and Sim entries are unavailable. Their 16-slot reference values are shown for context and do not form matched comparisons with these controls. Both transferred vectors in this table are rescaled to the target's own $\|\hat{b}\|_1$, making them magnitude diagnostics. The unrescaled coefficient-transfer results are reported separately in Table 19.

Table 21: Archived IL7R control diagnostics, with different sample sizes across metrics. ΔReward is the backbone-clustered paired difference against the corresponding 8-slot AbMPNN subset, and is the column comparable to Table 19; absolute values summarize each method separately. ΔG_{sep} is the median after one round of FastRelax (more negative is better); scRMSD is heavy-chain C_α RMSD after heavy-chain alignment (Å), based on 64 designs per method except the two norm-matched transfer variants (200). These are alphabetical subsets, not backbone-balanced samples. Sim uses up to 400 own designs and a fixed cross-target reference pool (Appendix D.6). Lower values indicate closer structural agreement or lower reference-sequence similarity, respectively. These columns are descriptive.

Method	**SC**	**ΔSASA**	**Hbonds**	ΔG_{sep}	**Reward**	**ΔReward**	**scRMSD**	**Sim**
AbMPNN, 16-slot mean	0.4070	1109.4	1.59	-16.5	9.937	—	1.835	0.235
AbMPNN, 8-slot paired subset	0.4040	1110.1	1.57	—	9.901	—	—	—
Shuffled	0.4036	1089.4	1.38	-15.7	9.803	-0.097	1.893	0.224
Own $\hat{b}$	0.4209	1131.6	1.51	-17.9	10.207	$+0.306$	1.753	0.194
Another antigen's (L_1-matched)	0.4249	1140.0	1.68	-17.5	10.299	$+0.398$	1.475	0.216
Pooled, rescaled to target norm	0.4338	1167.6	1.64	-18.3	10.527	$+0.627$	1.516	0.200

F.7 Magnitude-Matched Steering Controls

The transferred vectors in Table 19 retain the norms estimated on the source antigens. Because a weaker tilt would be an alternative explanation for any deficit, we also report them rescaled to the target's own $\|\hat{b}\|_1$ (Table 22). This is a magnitude diagnostic rather than a transfer result: the norm can only be known by fitting on the target, so these comparisons use target-specific information.

On HER2, the pooled vector gives a smaller reward gain than the target-fitted vector before rescaling. After matching the target norm, the direct difference is no longer significant ($p = 0.21$).

Table 22: Steering controls rescaled to the target's own $\|\hat{b}\|_1$, on the design set of Table 19. **Bold** marks the largest gain in each antigen column.

Steering vector	**HER2**		**IL7R**		**PD1**	
	Δ reward	p	Δ reward	p	Δ reward	p
Pooled, $\|b\|$-matched	$\mathbf{+0.940}$	4.4×10^{-18}	$\mathbf{+0.627}$	8.9×10^{-15}	$+0.382$	5.7×10^{-4}
Another's, $\|b\|$-matched	$+0.656$	3.1×10^{-8}	$+0.398$	8.3×10^{-6}	$\mathbf{+0.513}$	8.3×10^{-6}

F.8 Transfer to Other Decoders

Table 23 gives the numerical results behind Figure 6. It applies the pooled coefficient vector — estimated from AbMPNN decodes of the *other* antigens — unchanged to two policies that played no part in fitting it. The aggregate score improves in all six antigen–decoder combinations. Buried area accounts for 62–95% of the reward gain (Figure 11). None of the SC differences is significant, so these results establish reward transfer without a statistically resolved packing improvement.

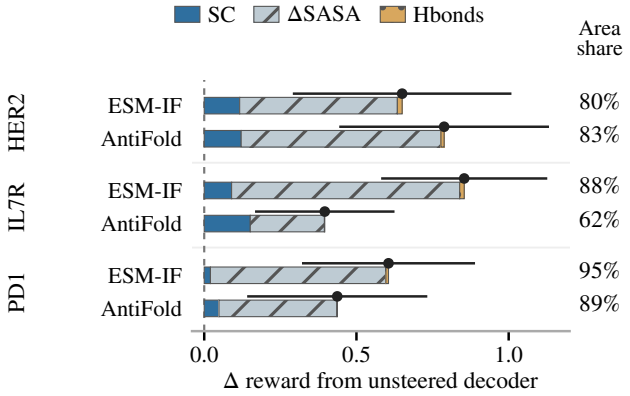


Figure 11: **Reward components in decoder transfer.** Stacked bars decompose the gains in Figure 6 into the weighted SC, ΔSASA, and H-bond terms. Dots and intervals show total mean gains and unadjusted 95% CIs from paired backbone means. Area share is the fraction of each gain contributed by ΔSASA. These component diagnostics are exploratory; no SC change is statistically resolved (Table 23).

Table 23: **Transfer to policies the bias was never fitted on.** Δ is paired against each policy's own unsteered output on identical backbones, with eight designs per backbone. HER2–AntiFold uses 127 backbones (1,016 pairs); the other five comparisons use 128 (1,024 pairs). Inference is clustered by backbone. **Bold** marks reward gains that pass the six-test correction ($\alpha' = 0.0083$); all six do. No SC gain is significant at $\alpha = 0.05$. Absolute means and paired differences use the same common design set. "SC share" is the fraction of the reward gain attributable to the shape-complementarity term, the remainder being predominantly interface area.

Antigen	Base policy	Base	+tilt	Δ reward	p	ΔSC	p(SC)	SC share
HER2	ESM-IF	12.576	13.226	**+0.650**	3.3×10^{-4}	+0.0090	0.20	18%
	AntiFold	11.930	12.718	**+0.788**	5.7×10^{-6}	+0.0094	0.22	15%
IL7R	ESM-IF	10.225	11.079	**+0.854**	4.7×10^{-10}	+0.0069	0.37	11%
	AntiFold	9.844	10.240	**+0.396**	5.6×10^{-4}	+0.0117	0.15	38%
PD1	ESM-IF	9.838	10.443	**+0.605**	2.3×10^{-5}	+0.0016	0.84	3%
	AntiFold	9.990	10.427	**+0.437**	3.3×10^{-3}	+0.0037	0.63	11%

F.9 Refolding Oracle Comparison

Reward repeatability. Figure 12 summarizes repeated PD1 evaluations from the archived score files. IgFold and Boltz-2 each have two predictions for 24 sequences from six backbones; RF2 has two predictions for 2,048 sequences from 128 backbones. We recompute reward with the same fixed weights and pair predictions by sequence identity. Repeat correlations are 0.994 for IgFold, 0.527 for RF2, and 0.323 for Boltz-2.

For paired rewards $R_i^{(1)}, R_i^{(2)}$, we summarize variability as $\mathrm{SD}(R^{(2)} - R^{(1)})/\sqrt{2}$, using the empirical SD with divisor n. This gives 0.121, 1.436, and 1.392 reward units, respectively. Interpreting it as per-run noise assumes independent repeats with comparable variances. The IgFold scores average three Rosetta repacks per prediction; the RF2 and Boltz-2 scores use one. Since both the sequence

sets and repacking budgets differ, these measurements characterize the archived protocols rather than a controlled ranking of the folding models. The archived PD1 split-half coefficient correlations, 0.620 with IgFold and 0.240 with RF2, provide a separate diagnostic of fitting stability.

Why fitting and evaluation use different protocols. Our IgFold pipeline aligns the antibody framework to the designed structure while keeping the antigen fixed. The resulting rewards are repeatable, but this procedure preserves the designed relative pose. Bulky residues can therefore increase buried area without improving SC. RF2 predicts the complex again and can change the antibody–antigen geometry, providing a stricter test of whether an interface-score gain survives refolding. Boltz-2 was explored as another complex predictor but is used for neither coefficient fitting nor the main held-out evaluation.

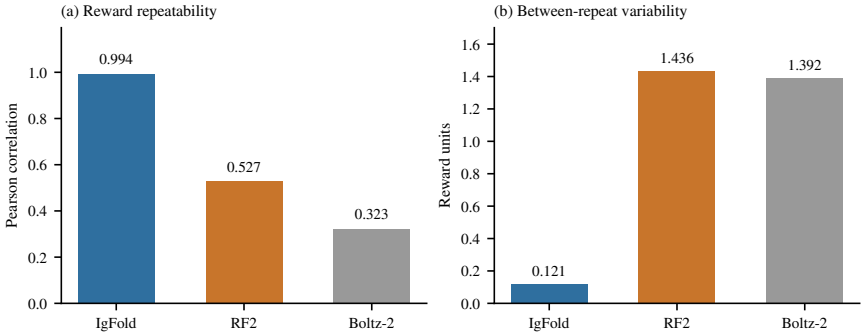


Figure 12: **Repeatability of archived PD1 reward evaluations.** (a) Correlation between two evaluations of each sequence. (b) SD of paired reward differences divided by $\sqrt{2}$. IgFold and Boltz-2 use 24 sequences each; RF2 uses 2,048. The cohorts and Rosetta repacking budgets differ as described in the text. These are descriptive protocol diagnostics, not confidence intervals or a matched comparison of folding models.

F.10 Scope and Limitations

Surrogate validity. All results are computational. The SKEMPI-fitted interface score has no held-out predictive validation, and applying it to multi-residue CDR designs on generated backbones introduces distribution shift. Results also depend on backbone quality and the structural oracle. Experimental binding measurements are needed to determine whether score improvements translate to affinity.

Restricted expressivity and partial transfer. PACT captures composition but lacks position- or context-specific coefficients. The per-loop extension does not significantly outperform the shared tilt on HER2 ($+0.026$, $p = 0.41$; Appendix F.4), which is insufficient to rule out benefits from richer families. A pooled vector improves the aggregate score on all three antigens, with statistically resolved SC gains on two, and its HER2 deficit is sensitive to rescaling with the target's norm. Direct comparisons with the volume-only control are inconclusive on HER2 and IL7R, while favoring the fitted vector on PD1. Together these results leave the balance between generic composition effects and target-specific steering unresolved beyond the tested settings.

Scope of the optimization diagnosis. The residual decomposition applies at any iterate, and controlled checkpoint probes support the local linearization for SGD and AdamW (Appendix E.5). These probes cover one advantage-regression run per antigen, with fixed candidates and decoding order. They do not reconstruct historical minibatches or establish a causal account of the full training trajectory. The conditional path bound in Equation (21) has not been numerically certified. Only two full-parameter objectives have refolded performance evaluations, and filtered SFT is evaluated after near-exact imitation rather than after held-out reward-based checkpoint selection. The other four objectives have limited diagnostic coverage, so the experiments do not rule out better full-parameter training procedures.