DIRECT OR Δ? UNDERSTANDING TWO ROUTES TO MUTANT BINDING AFFINITY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting how protein mutations alter ligand binding is crucial for understanding drug response and identifying compounds effective across variants. In modern biomolecular modeling, increasingly capable models coexist with comparatively limited mutation-specific affinity data, making the formulation of the prediction target itself an important modeling choice. Mutant affinity can be predicted either directly or through its change relative to the wild type (WT), yet it remains unclear when one route should be preferred over the other. We develop a statistical framework for comparing the two routes, Direct and Delta (Δ), under the same input. The analysis shows when WT–mutant pairing can simplify affinity-change prediction by canceling shared variation, how this advantage depends on training size, and how interacting prediction errors affect the reconstructed mutant affinity. A linear-Gaussian analysis further characterizes conditions under which the preferred route can change with sample size. Controlled simulations reproduce the predicted finite-sample behavior, while experiments on PLATINUM and DAVIS-Complete reveal distinct real-data regimes: the Δ-route advantage narrows as Direct prediction improves on PLATINUM, whereas sample-specific affinity-change prediction becomes increasingly informative with additional data on DAVIS-Complete. These results highlight the framework as a tool both explanatory and practical for modeling. *Our code will be publicly available.*

1 INTRODUCTION

Protein mutations can alter binding affinity for small molecules and potentially affect drug response. A well-known example is the T315I substitution in BCR-ABL, which interferes with imatinib binding and contributes to drug resistance in chronic myeloid leukemia (Gorre et al., 2001). A more recent example is the G23 deletion in SARS-CoV-2 main protease, which markedly weakens ensitrelvir binding and confers antiviral resistance (Ma et al., 2026). Predicting mutant–ligand binding affinity is therefore important both for understanding mutation-associated changes in drug response and for identifying compounds that remain effective across protein variants.

For a mutant protein, the corresponding unmutated protein is referred to as the wild type (WT), providing a natural reference for two formulations of mutant-affinity prediction as shown in Figure 1: the Direct-route and the Δ-route. The Direct-route predicts the mutant affinity itself; for example, DeepMutDTA retains absolute mutant affinity as the prediction target (Hu et al., 2026). In contrast, the Δ-route predicts the mutation-induced change in affinity relative to WT. Methods such as MPLBind model the effect of mutations on protein-ligand affinity, while DDMuffin explicitly predicts mutation-induced affinity changes in the form of $\Delta\Delta G$ (Jin & Li, 2025; Zhou et al., 2025).

Today, biomolecular models continue to grow in capacity and pretraining scale, while mutation-specific binding-affinity data remain comparatively limited (Gelman et al., 2025). This makes the choice of prediction target important for learning efficiently from the available data and motivates our central research question:

When and why should mutant binding affinity be predicted directly or through Δ?

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.

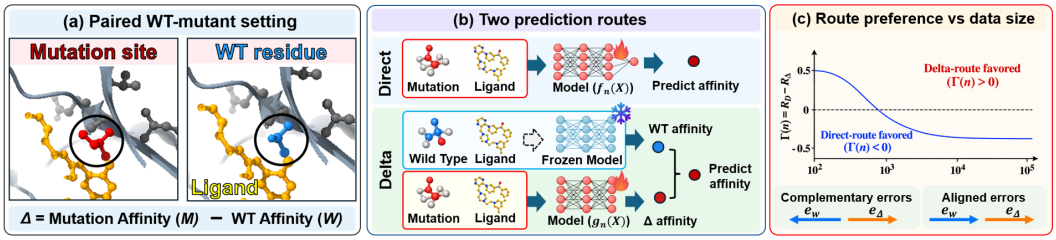

Figure 1: **Direct and Δ routes for mutant binding-affinity prediction. (a)** A paired WT–mutant system defines the affinity change $\Delta = M - W$. **(b)** Both routes use the same input information: Direct learns mutant affinity M, while the Δ-route learns the affinity change and reconstructs M with a WT-affinity reference. **(c)** Route advantage can vary with training size, while complementary or aligned component errors can respectively reduce or increase the reconstructed error.

In this work, we develop a statistical framework for comparing the Direct and Δ-routes. We characterize their target difficulty through

$$\Omega_\Delta = \Omega_M + \Omega_W - 2\Omega_{MW},$$

showing when shared WT–mutant variation makes Δ easier to learn. We then analyze how this advantage translates to final mutant-affinity prediction when Δ is combined with a frozen WT reference. Route preference therefore depends on target difficulty, WT-reference error, and their error interaction across data regimes. Our analysis leads to three practical takeaways:

When	Route preference can shift with data scale. The Δ-route may be advantageous in small-data regimes, while the Direct-route can become preferable as training data grow.
Why	WT and mutant affinities share much of the same biological and experimental context when acquiring them, allowing shared variation to cancel with the Δ-route.
What	Modeling Δ is valuable when mutation-specific affinity changes can be learned beyond the average mutation effect.

Controlled simulations reproduce the predicted route crossover, showing that the Δ route can benefit from shared WT–mutant variation in limited-data regimes, whereas Direct prediction can catch up and become preferable as training data increase. Experiments on PLATINUM (Pires et al., 2015) and DAVIS-Complete (Wu et al., 2025) further reveal two distinct real-data regimes. On PLATINUM, the advantage of the Δ route narrows as Direct prediction improves with additional training data, demonstrating the finite-sample route dependence predicted by our theoretical analysis. On DAVIS-Complete, the route advantage shows no clear monotonic trend, while the ability to predict sample-specific affinity changes improves steadily with data. Taken together, these results highlight the framework as both an explanatory and practical tool: it clarifies how the balance between shared WT–mutant structure and learnable mutation-specific variation changes with data scale, thereby guiding the route choice.

Our main contributions are:

- We establish a statistical framework for comparing Direct and Δ-routes through the shared structure of paired WT–mutant affinities.
- We derive a finite-sample route criterion that characterizes how training scale and error interaction shape route preference.
- We validate the theoretical predictions through controlled simulations and experiments on PLATINUM and DAVIS-Complete, confirming data-scaling behaviors of the two routes.

2 Methodology

2.1 Affinity Prediction Routes

Let X contain all non-label information available, such as the WT protein, mutation descriptor and ligand. Let W and M denote the WT and mutant affinities on the same transformed scale, and we

assume they are with finite second moments. The biological mutation effect is defined as

$$\Delta = M - W. \tag{1}$$

We consider two routes for predicting mutant affinity M from X. The *direct-route* estimates M as

$$\widehat{M}_{\mathrm{D}} = f(X), \tag{2}$$

whereas the Δ-*route* obtains mutant affinity as

$$\widehat{M}_{\Delta} = a(X) + g(X), \tag{3}$$

where $a(X)$ is a frozen WT-affinity predictor and $g(X)$ is trained to predict the paired mutation-effect label Δ. The paired WT affinity W is required during training to construct $\Delta = M - W$, while the true W is not used at inference.

2.2 Statistical Basis of the Two Routes

The two routes differ in their prediction targets. To characterize this difference, define

$$m(X) = \mathbb{E}[M \mid X], \qquad w(X) = \mathbb{E}[W \mid X], \tag{4}$$

with residuals

$$\xi_M = M - m(X), \qquad \xi_W = W - w(X). \tag{5}$$

By construction, $\mathbb{E}[\xi_M \mid X] = \mathbb{E}[\xi_W \mid X] = 0$. And,

$$\Delta = \underbrace{m(X) - w(X)}_{X\text{-predictable component } \delta(X)} + \underbrace{\xi_M - \xi_W}_{\text{MSE-irreducible residual}}, \tag{6}$$

where $\delta(X) = \mathbb{E}[\Delta \mid X]$.

To characterize the residual difficulty of the two targets, define

$$\begin{aligned} \Omega_M &:= \mathrm{Var}(\xi_M), \\ \Omega_W &:= \mathrm{Var}(\xi_W), \\ \Omega_{MW} &:= \mathrm{Cov}(\xi_M, \xi_W). \end{aligned} \tag{7}$$

The Direct target M has residual variation Ω_M, whereas the Δ target has

$$\Omega_{\Delta} := \mathrm{Var}(\xi_M - \xi_W) = \Omega_M + \Omega_W - 2\Omega_{MW}. \tag{8}$$

Hence, the Δ-route reduces residual variation relative to direct prediction precisely when

$$\Omega_{\Delta} < \Omega_M \quad \Longleftrightarrow \quad 2\Omega_{MW} > \Omega_W. \tag{9}$$

This provides a preliminary target-level indication of when and why one route may be more favorable than the other.

2.3 Prediction Error of the Two Routes

We now compare the two routes in terms of their final mutant-affinity prediction error. At training budget n, let f_n and g_n denote the fitted predictors for M and Δ, respectively.

For any route's prediction $\widehat{M}(X)$, recall that $M = m(X) + \xi_M$. Therefore,

$$\widehat{M}(X) - M = \big(\widehat{M}(X) - m(X)\big) - \xi_M. \tag{10}$$

Since $\mathbb{E}[\xi_M \mid X] = 0$, the two terms are orthogonal, giving

$$\mathbb{E}[(\widehat{M}(X) - M)^2] = \Omega_M + \mathbb{E}[(\widehat{M}(X) - m(X))^2]. \tag{11}$$

Thus, both routes share the same irreducible mutant-affinity variation Ω_M, and their difference depends on how closely their fitted predictions approximate $m(X)$.

For the Δ route, we have

$$a(X) + g_n(X) - m(X) = \big(a(X) - w(X)\big) + \big(g_n(X) - \delta(X)\big). \tag{12}$$

Define the frozen-WT error

$$\epsilon_a := \mathbb{E}[(a(X) - w(X))^2]. \tag{13}$$

Theorem 1 (Prediction-error) *The test-population mean squared errors of the two routes are*

$$R_{\mathrm{D}}(n) := \mathbb{E}[(f_n(X) - M)^2] = \Omega_M + \mathbb{E}[(f_n(X) - m(X))^2], \tag{14}$$
$$R_{\Delta}(n) := \mathbb{E}[(a(X) + g_n(X) - M)^2] = \Omega_M + \mathbb{E}[(a(X) + g_n(X) - m(X))^2].$$

Consequently, the route advantage is

$$\Gamma(n) := R_{\mathrm{D}}(n) - R_{\Delta}(n) = \mathbb{E}[(f_n(X) - m(X))^2] - \mathbb{E}[(g_n(X) - \delta(X))^2] - \epsilon_a - \kappa_a(n). \tag{15}$$

where

$$\epsilon_a := \mathbb{E}[(a(X) - w(X))^2], \quad \kappa_a(n) := 2\mathbb{E}[(a(X) - w(X))(g_n(X) - \delta(X))]\,. \tag{16}$$

The Δ route has lower final mutant-affinity MSE if and only if $\Gamma(n) > 0$.

Remark 1 (Error interaction) *A positive $\kappa_a(n)$ means that the frozen-WT and Δ prediction errors tend to reinforce each other, whereas a negative $\kappa_a(n)$ indicates partial cancellation between them.*

Remark 2 (Large-sample behavior) *Suppose both fitted predictors are L_2-consistent for their population targets,*

$$\mathbb{E}[(f_n(X) - m(X))^2] \to 0, \qquad \mathbb{E}[(g_n(X) - \delta(X))^2] \to 0,$$

as $n \to \infty$, with $\epsilon_a < \infty$. Then $\kappa_a(n) \to 0$ by the Cauchy–Schwarz inequality, and

$$\Gamma(n) \to -\epsilon_a.$$

Thus, if $\epsilon_a > 0$, any finite-sample advantage of the Δ route vanishes asymptotically; if $\epsilon_a = 0$, the two routes become asymptotically equivalent.

2.4 Linear–Gaussian special case.

To make the dependence of the learning errors on training size regarding $\Gamma(n)$ explicit, we consider a correctly specified linear–Gaussian setting in which both targets are fitted by ordinary least squares using the same p regressors. Let $\mathcal{D}_n$ denote the random training sample of size n, and assume a Gaussian random design, homoscedastic residual variation, and $n > p+1$. Under these assumptions, averaging over random training samples gives

$$\mathbb{E}_{\mathcal{D}_n,X}\left[(f_n(X) - m(X))^2\right] = c(n)\Omega_M, \quad \mathbb{E}_{\mathcal{D}_n,X}\left[(g_n(X) - \delta(X))^2\right] = c(n)\Omega_{\Delta}, \tag{17}$$

where

$$c(n) := \frac{p}{n - p - 1}. \tag{18}$$

As ordinary least squares is unbiased under the stated specification and the WT predictor is fixed,

$$\mathbb{E}_{\mathcal{D}_n}[\kappa_a(n)] = 0. \tag{19}$$

Therefore, the expected prediction errors of the two routes are

$$\mathbb{E}_{\mathcal{D}_n}[R_{\mathrm{D}}(n)] = \Omega_M + c(n)\Omega_M, \tag{20}$$
$$\mathbb{E}_{\mathcal{D}_n}[R_{\Delta}(n)] = \Omega_M + \epsilon_a + c(n)\Omega_{\Delta}. \tag{21}$$

Their expected route advantage is consequently

$$\mathbb{E}_{\mathcal{D}_n}[\Gamma(n)] = c(n)(\Omega_M - \Omega_{\Delta}) - \epsilon_a = c(n)(2\Omega_{MW} - \Omega_W) - \epsilon_a. \tag{22}$$

The first form shows that the finite-sample advantage depends on the difference in residual variation between the two prediction targets. The second makes the underlying WT–mutant dependence explicit: stronger positive residual covariance Ω_{MW} increases the advantage of differencing.

For the route comparison, this effect is summarized by the residual variation gap $\Omega_M - \Omega_{\Delta}$. A crossover is possible when this gap is positive and $\epsilon_a > 0$: the former gives the Δ-route a finite-sample estimation advantage, while the latter introduces a persistent penalty.

$$n^* = p + 1 + \frac{p(\Omega_M - \Omega_{\Delta})}{\epsilon_a}. \tag{23}$$

Thus, within this special case, the Δ-route is favored below n^* and Direct above it.

2.5 Empirical Guidance for Route Selection

To translate the theoretical analysis into practical evaluation, we organize three empirical questions and define corresponding measures for subsequent analysis.

When is the Δ-route advantageous? The theoretical analysis predicts that route preference can depend on training size. We therefore estimate the paired route advantage on the held-out examples,

$$\widehat{\Gamma}(n) := \frac{1}{N}\sum_{i=1}^{N}\left[(f_n(X_i) - M_i)^2 - (a(X_i) + g_n(X_i) - M_i)^2\right]. \tag{24}$$

We track $\widehat{\Gamma}(n)$ across training budgets; as the training size increases, the theory predicts that $\widehat{\Gamma}(n)$ may decrease as the finite-sample advantage of the Δ-route weakens.

Why does the Δ route gain or lose its advantage? We examine two perspectives. First, we examine whether shared WT–mutant variation is reduced by differencing. We compare the held-out residual variation of mutant affinity and Δ through $\widehat{\Omega}_M - \widehat{\Omega}_\Delta$. A positive gap indicates smaller residual variation for Δ under the fitted models. Since these residuals also contain conditional-mean estimation error, the gap provides a model-dependent diagnostic of the mechanism discussed in Eq. (9).

Second, we examine how prediction errors combine when the mutant affinity is reconstructed. For each held-out example, define $e_{W,i} := a(X_i) - W_i$, $e_{\Delta,i} := g_n(X_i) - \Delta_i$. Since $M_i = W_i + \Delta_i$,

$$a(X_i) + g_n(X_i) - M_i = e_{W,i} + e_{\Delta,i}. \tag{25}$$

Therefore, the empirical endpoint risk of the Δ route is

$$\widehat{R}_\Delta(n) = \frac{1}{N}\sum_{i=1}^{N} e_{W,i}^2 + \frac{1}{N}\sum_{i=1}^{N} e_{\Delta,i}^2 + \frac{2}{N}\sum_{i=1}^{N} e_{W,i}e_{\Delta,i}. \tag{26}$$

The first two terms quantify WT-reference and Δ-prediction errors, while the final term captures whether these errors reinforce or partially cancel each other. A negative interaction reduces the reconstructed mutant-affinity error, motivating joint evaluation of the WT and Δ predictors.

What does the Δ model learn beyond the average effect? A lower final mutant-affinity error does not necessarily mean that $g_n(X)$ distinguishes mutation effects across samples. For example, a constant predictor may still provide a useful average correction. We therefore compare $g_n(X)$ with the training-set mean effect $\bar{\Delta}_{\text{train}}$:

$$\text{Gain}_\Delta(n) := 1 - \frac{\text{MSE}(g_n(X), \Delta)}{\text{MSE}(\bar{\Delta}_{\text{train}}, \Delta)}, \tag{27}$$

where $\bar{\Delta}_{\text{train}} = n^{-1}\sum_{i=1}^{n}\Delta_i$ is the mean mutation effect in the training set. A positive $\text{Gain}_\Delta(n)$ indicates that $g_n(X)$ outperforms the constant mean-effect baseline, whereas a non-positive value indicates no improvement over this baseline. Thus, $\text{Gain}_\Delta(n)$ measures whether modeling sample-specific variation in Δ provides predictive value beyond a mean correction.

3 Experiments

3.1 Experimental Settings

Datasets. We use 370 paired WT–mutant affinity records from PLATINUM and 831 from DAVIS-Complete, the latter spanning 271 WT–ligand context groups. Splits are disjoint by WT–ligand context but may share ligand identities. The train/validation/development splits contain 258/56/56 records for PLATINUM and 581/125/125 for DAVIS-Complete. Appendix A.1 provides details.

Synthetic setup. We additionally use a controlled linear–Gaussian simulation to isolate the effects of training size, frozen-anchor bias, and residual WT–mutant coupling. Both routes use the same 20-dimensional Gaussian inputs and training examples, and fit 20 coefficients by unregularized ordinary least squares without an intercept. Each experimental setting is averaged over 1,000 independent Monte Carlo replications. The full data-generating process is detailed in Appendix A.2.

Models. The Direct and Δ learners share a shallow late-fusion MLP with identical frozen protein, mutation-site, and ligand representations. They differ only in their regression targets. The Δ route additionally uses a separately trained WT predictor, frozen across training sizes and subset chains within each seed. We include a matched anchor-residual learner as an additional control. Model architectures, regression targets, and WT predictor training are detailed in Appendix A.3.

Training protocol. Experiments are repeated over five random seeds. The Direct and Δ learners are trained with AdamW using a learning rate of 10^{-3}, weight decay of 10^{-4}, and batch size 32 for up to 200 epochs, with early stopping after 25 epochs without validation improvement. Checkpoints are selected by validation MSE on the route's target. Details of training-size subsampling and uncertainty estimation are provided in Appendix A.4.

3.2 Route Preference under Controlled Conditions

We use controlled simulations to isolate the three quantities in Eq. (22) governing finite-sample route preference: training size n, WT-reference error ϵ_a, and residual WT–mutant covariance Ω_{MW}. We choose a common baseline near the theoretical route boundary ($n = 96$, $\epsilon_a = 0.15$, and $\Omega_{MW} = 0.8$), and vary one quantity at a time while keeping the other two fixed. Overall, Figure 2 shows close agreement between the theoretical predictions and simulation results.

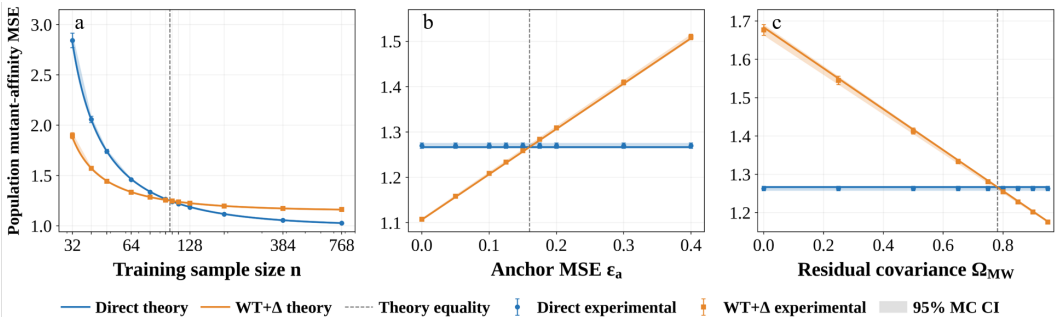

Figure 2: Controlled comparison of the Direct and Δ routes in the linear–Gaussian setting. We vary **(a)** training sample size, **(b)** WT-reference error ϵ_a, and **(c)** residual WT–mutant covariance Ω_{MW}, one at a time. Solid curves are theoretical expected mutant-affinity MSE, markers show means over 1,000 independent simulations, and vertical dashed lines mark the theoretical route-equality points.

Training size. Varying the training size alone produces the predicted route crossover as shown in Figure 2(a): the Δ route is preferred with fewer observations, whereas Direct becomes preferable as the sample size grows. The theoretical crossover occurs at $n^* = 101$, while the simulation gives $\hat{n}^* = 100.384$ with a 95% paired bootstrap interval of $[96.960, 103.848]$. It confirms that the advantage of learning the difference target can be finite-sample: as training size grows, the finite-sample benefit of the difference target diminishes, whereas the WT-reference error remains fixed.

WT-reference error. Figure 2(b) isolates the effect of WT-reference error. As ϵ_a increases, the Δ-route error increases, with the two routes becoming equal at $\epsilon_a^* = 0.16$. It shows WT-reference quality can determine whether the advantage of learning Δ carries over to mutant-affinity prediction.

WT–mutant residual coupling. Figure 2(c) isolates the effect of WT–mutant residual coupling. As Ω_{MW} increases, more shared residual variation is canceled by differencing, reducing Ω_Δ (see Eq. (8)) from 2 to 0.1 over the simulated range. The resulting reduction in the Δ-route error leads to a theoretical crossover at $\Omega_{MW}^* = 0.78125$. Thus, stronger WT–mutant residual coupling can make the difference target sufficiently easier to learn for the Δ route to become preferable.

3.3 Route Behavior on Real-World Data

We next examine how the two routes behave on PLATINUM and DAVIS-Complete, focusing on three empirical questions defined in Section 2.5.

Table 1: Training-size behavior on PLATINUM and DAVIS-Complete. Values in brackets are pointwise 95% confidence intervals. The final row reports slopes *w.r.t.* log training size where evaluated.

Budget	PLATINUM $\widehat{\Gamma}$		$\widehat{\Omega}_M - \widehat{\Omega}_\Delta$		Gain_Δ		DAVIS $\widehat{\Gamma}$		$\widehat{\Omega}_M - \widehat{\Omega}_\Delta$		Gain_Δ	
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI
20%	3.485	[0.84, 7.79]	4.421	[1.48, 9.52]	-0.048	[−0.12, 0.03]	0.255	[−0.61, 1.27]	1.702	[0.56, 3.07]	0.125	[0.07, 0.19]
40%	1.946	[0.57, 4.14]	2.810	[1.15, 5.70]	-0.057	[−0.13, 0.03]	0.324	[0.05, 0.57]	1.721	[0.78, 2.71]	0.196	[0.09, 0.29]
60%	1.510	[0.41, 3.22]	2.271	[0.90, 4.33]	-0.111	[−0.28, −0.01]	0.267	[0.04, 0.54]	1.558	[0.63, 2.62]	0.267	[0.14, 0.38]
80%	1.207	[0.04, 2.98]	1.901	[0.63, 3.91]	-0.109	[−0.23, 0.01]	0.382	[0.07, 0.71]	1.564	[0.59, 2.72]	0.304	[0.15, 0.44]
100%	0.689	[−0.12, 2.08]	1.256	[−0.27, 3.37]	-0.174	[−0.41, 0.01]	0.357	[−0.09, 0.83]	1.788	[0.66, 3.12]	0.353	[0.14, 0.52]
Slope	-1.678	[−3.78, −0.38]	–	–	-0.072	[−0.173, 0.002]	0.064	[−0.79, 0.77]	–	–	0.139	[0.04, 0.23]

Table 2: Full-budget results on PLATINUM and DAVIS-Complete. Endpoint MSE is reported as mean ± standard deviation over five matched seeds. Gain_Δ measures improvement in Δ prediction over the training-set mean-effect baseline.

(a) Mutant-affinity MSE ↓

Route	PLATINUM	DAVIS-Complete
Direct	2.7916 ± 0.1155	2.7852 ± 0.0449
WT reference	2.1796 ± 0.2102	2.5161 ± 0.3606
WT + mean Δ	$\mathbf{1.9925 \pm 0.1455}$	2.6318 ± 0.4204
Δ route	2.1026 ± 0.1376	$\mathbf{2.4284 \pm 0.3343}$
WT + residual	2.4589 ± 0.1349	2.6226 ± 0.1121

(b) Δ prediction gain ↑

Predictor	PLATINUM	DAVIS-Complete
Mean Δ	0	0
Learned Δ	-0.1737	$\mathbf{0.3525}$

Obs 1–When: Route preference across training sizes. Table 1 reports the training-size analysis for both datasets; here we focus on the route advantage $\widehat{\Gamma}(n)$. Figure 3 visualizes the corresponding training-size trajectories. Detailed experimental settings are provided in Appendix A.4.

On PLATINUM, the estimated route advantage in Eq. (24) decreases from $\widehat{\Gamma} = 3.485$ at the 20% training budget to 0.689 at the full training size. A linear regression of $\widehat{\Gamma}$ on log training size gives $\beta_\Gamma = -1.678$ (95% CI: $[-3.78, -0.38]$), supporting a decreasing trend. This narrowing is driven mainly by Direct, whose endpoint MSE decreases from 5.562 to 2.792, while the Δ-route remains approximately stable at 2.077–2.103. On DAVIS-Complete, $\widehat{\Gamma}$ changes from 0.255 to 0.357 across training budgets, with no resolved monotonic trend ($\beta_\Gamma = 0.064$, 95% CI: $[-0.79, 0.77]$). Thus, no increasing or decreasing trend is statistically resolved over the evaluated training sizes.

PLATINUM mirrors the behavior observed in the controlled simulation. In the linear–Gaussian setting, the finite-sample advantage of an easier Δ target decreases with training size when the remaining factors are fixed. In real-world settings, however, other quantities governing $\Gamma(n)$ may also evolve with training size, so a monotonic decrease is not guaranteed. The following observations examine the mechanisms behind these contrasting behaviors. The key takeaway for Obs 1 is:

> **Takeaway 1.** *Route preference depends on data scale.* The Δ-route can be more competitive with limited training data, whereas the Direct route can catch up or become preferable as the training set grows. The exact trend also depends on target difficulty, WT-reference accuracy, and reconstruction-error interaction.

Obs 2–Why: Shared variation and error cancellation. We first examine the target-level mechanism predicted by the theory. Across training budgets, the point estimates of $\widehat{\Omega}_M - \widehat{\Omega}_\Delta$ are positive on both datasets (Table 1). On DAVIS-Complete, the 95% confidence intervals remain above zero throughout, whereas on PLATINUM the full-budget interval includes zero. These diagnostics are consistent with shared-variation cancellation, although they do not separate irreducible variation from conditional-mean estimation error.

We next examine the reconstruction stage using the held-out decomposition in Eq. (26). At the full training size, the WT-reference MSE, Δ-prediction MSE, and interaction term are 2.7780, 1.0637, and -1.7391 on PLATINUM, yielding $\widehat{R}_\Delta = 2.1026$. The corresponding values on DAVIS-Complete are 2.8187, 0.7079, and -1.0982, yielding $\widehat{R}_\Delta = 2.4284$. In both datasets,

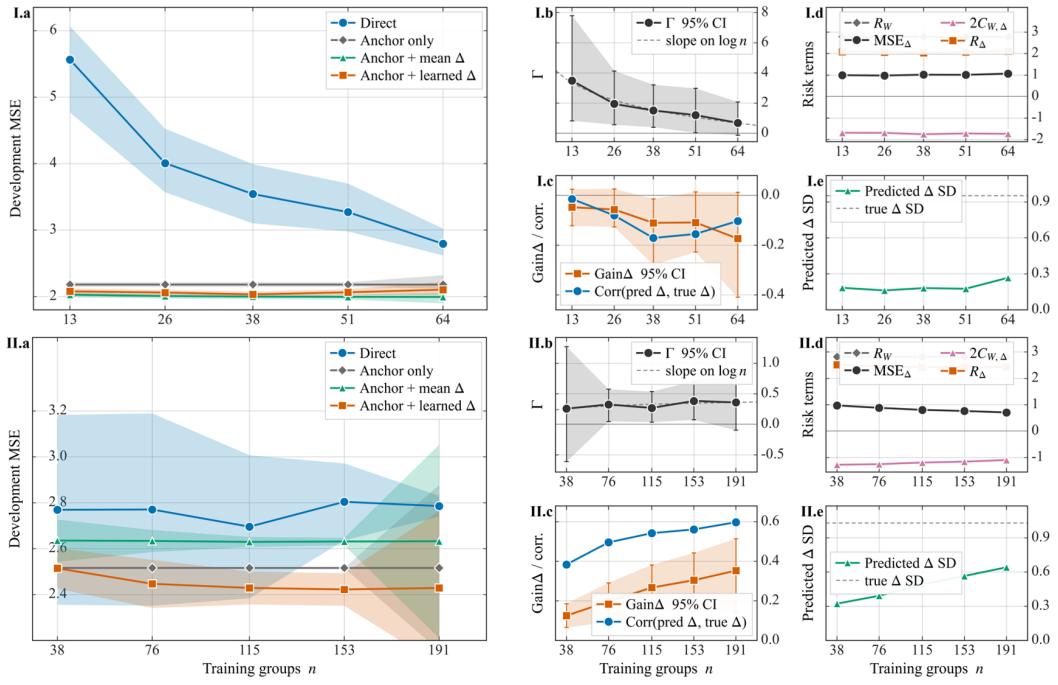

Figure 3: Training-size analysis on PLATINUM (I) and DAVIS-Complete (II): (a) shows development-set mutant-affinity MSE; the (b) shows the paired route advantage Γ; (c) reports mutation-effect gain and the correlation between predicted and observed effects; (d) decomposes the Δ-route endpoint MSE into the WT-reference error, Δ-prediction error, and their interaction; and (e) compares the standard deviations of predicted and observed mutation effects.

the negative interaction offsets a substantial portion of the two MSE terms and reduces the reconstructed mutant-affinity error. Because the observed WT affinity enters $e_W = a(X) - W$ and $e_\Delta = g_n(X) - (M - W)$ with opposite signs, part of this cancellation can arise structurally from the paired WT–mutant construction. A detailed attribution is provided in Appendix A.5.

These findings also help explain the contrasting training-size behaviors of $\widehat{\Gamma}(n)$ observed in Obs. 1. On PLATINUM, the estimated residual gap $\widehat{\Omega}_M - \widehat{\Omega}_\Delta$ decreases from 4.421 to 1.256 as training size grows, while the Δ-prediction MSE and interaction term change little. Together with the substantial improvement of Direct, these changes accompany the decrease in $\widehat{\Gamma}(n)$. On DAVIS-Complete, the estimated residual gap remains positive across training budgets, while the Δ-prediction error continues to decrease with additional data. These pattern is consistent with the lack of a resolved decreasing trend in $\widehat{\Gamma}(n)$. The key takeaway for Obs 2 is:

> **Takeaway 2.** *Shared WT–mutant structure can favor the Δ route.* Differencing can reduce shared variation between WT and mutant affinities, while component errors can further cancel during affinity reconstruction.

Obs 3–What: Prediction beyond the mean effect. We finally examine whether the learned Δ model captures sample-specific affinity changes beyond the average mutation effect, as defined by Gain_Δ in Eq. (27). On PLATINUM, the full-budget gain is -0.174 (95% CI: $[-0.41, 0.01]$), and no increasing trend is resolved across training budgets ($\beta_{\mathrm{Gain}} = -0.072$, 95% CI: $[-0.173, 0.002]$). Consistently, the mean-effect correction achieves a lower endpoint MSE than the learned Δ correction (1.993 *vs.* 2.103). Thus, on PLATINUM, the learned sample-specific Δ does not outperform simply applying the average mutation effect.

DAVIS-Complete shows a different regime. Gain_Δ increases from 0.125 to 0.353 as training size grows, with pointwise 95% confidence intervals above zero at every evaluated budget and a positive

scaling trend ($\beta_{\mathrm{Gain}} = 0.139$, 95% CI: $[0.04, 0.22]$). Over the same range, the Δ-prediction MSE decreases from 0.973 to 0.708, while the prediction–target correlation increases from 0.384 to 0.597. At the full budget, the learned Δ correction also improves endpoint MSE over the mean-effect correction (2.428 *vs.* 2.632). These results indicate that additional data make sample-specific affinity changes increasingly predictable on DAVIS-Complete.

The two datasets therefore distinguish the value of the Δ route from the value of learning heterogeneous mutation effects. On PLATINUM, a sample-specific Δ model does not outperform the average-effect baseline, whereas on DAVIS-Complete it provides increasing predictive value as more data become available. Here is the key takeaway:

Takeaway 3. *The Δ route is valuable when mutation-specific affinity changes are learnable beyond the average effect.* A positive Gain_{Δ} indicates that modeling sample-specific variation provides predictive value beyond a common mean correction.

4 Related Work

Mutation-effect prediction for protein–ligand binding. Predicting how protein mutations alter ligand binding has been studied using structure-based, physics-based, and ML approaches. PLATINUM provides a widely used collection of experimentally measured mutation-induced affinity changes in structurally characterized protein–ligand complexes (Pires et al., 2015). Methods such as mCSM-lig use graph-based structural signatures to predict changes in protein–small-molecule affinity (Pires et al., 2016), while PremPLI combines sequence-, structure-, and ligand-derived features to estimate mutation-induced binding changes (Sun et al., 2021). More recent evaluations have further examined the robustness of such mutation-aware predictors when experimental complex structures are replaced by computationally predicted structures (Pan et al., 2026). These studies focus on affinity-change prediction, leaving open when reconstruction from a WT reference and a predicted change is statistically preferable to direct mutant-affinity prediction.

Difference and Δ learning. Learning differences rather than absolute targets has a broader history in molecular machine learning. Δ-machine learning combines a lower-cost baseline with a learned correction to approximate a higher-fidelity target (Ramakrishnan et al., 2015). Pairwise Difference Regression similarly learns property differences between related examples and uses them to improve absolute-property prediction and uncertainty estimation (Tynes et al., 2021). DeepDelta extends this idea to molecular optimization by directly learning property changes between molecular pairs and reports advantages over conventional absolute-property prediction on several ADMET tasks (Fralish et al., 2023). These results motivate the intuition that differencing can remove shared variation and yield a statistically simpler target. However, a smaller or more predictable difference target does not by itself guarantee a better reconstructed endpoint: the quality of the baseline or anchor and the interaction between its error and the learned correction also matter.

5 Conclusion

We compared Direct prediction of mutant affinity with a Δ-route that reconstructs mutant affinity from a WT reference and a predicted mutation-induced change. Our results show that route preference depends on data scale and on the statistical structure of paired WT–mutant measurements. The key insight is not whether the Δ route outperforms Direct, but *whether learning a sample-specific Δ adds information beyond a constant mutation effect*. On PLATINUM, the route gap narrows as Direct improves, while the current Δ predictor shows no supported improvement over the mean-effect baseline. On DAVIS-Complete, this additional signal emerges with more data, as sample-specific Δ prediction increasingly exceeds the mean-effect control. Thus, a Δ formulation is most informative when mutation-specific variation itself is learnable.

AI USE STATEMENT

The research question and core ideas originated with the authors. We used ChatGPT for language polishing, LaTeX formatting, and assistance with reviewing derivations. AI tools did not determine the research objectives or make final scientific decisions. The authors reviewed and verified all retained AI-assisted material and take full responsibility for the final manuscript.

ETHICS STATEMENT

This study uses synthetic data and publicly available protein–ligand affinity datasets. It involves no human participants, new animal experiments, or personally identifiable information. We identified no specific ethical concerns associated with this work.

REPRODUCIBILITY STATEMENT

Experimental settings are described in Section 3.1, with further details on data preparation, model architectures and uncertainty estimation in Appendices A.1, A.3 and A.4, respectively. The source code will be made publicly available to facilitate reproduction of our experiments.

ACKNOWLEDGMENTS

REFERENCES

Zachary Fralish, Ashley Chen, Paul Skaluba, and Daniel Reker. Deepdelta: predicting admet improvements of molecular derivatives with deep learning. *Journal of cheminformatics*, 15(1):101, 2023.

Sam Gelman, Bryce Johnson, Chase R Freschlin, Arnav Sharma, Sameer D'Costa, John Peters, Anthony Gitter, and Philip A Romero. Biophysics-based protein language models for protein engineering. *Nature Methods*, 22(9):1868–1879, 2025.

Mercedes E Gorre, Mansoor Mohammed, Katharine Ellwood, Nicholas Hsu, Ron Paquette, P Nagesh Rao, and Charles L Sawyers. Clinical resistance to sti-571 cancer therapy caused by bcr-abl gene mutation or amplification. *Science*, 293(5531):876–880, 2001.

Xiaowen Hu, Pan Zhang, Shangqian Wu, Hao Sun, Minwei Li, Sophia Tsoka, Zizhang Sheng, and Lei Deng. Deep contrastive learning for high-throughput prediction of drug resistance mutations from sequences. *Advanced Science*, pp. e14899, 2026.

Jiahuan Jin and Jie Li. Mplbind: Predicting the effect of binding site point mutations on protein–ligand binding affinity using protein large language models. *Journal of Chemical Information and Modeling*, 65(16):8720–8729, 2025.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Yao Ma, Chengjin Ye, R Rahisuddin, Sara H Mahmoud, Anastasija Cupic, Ahmed Magdy Khalil, Esteban Castro, Nathaniel Jackson, Mahmoud Bayoumi, Shailee Arya, et al. Identification and characterization of a sars-cov-2 mpro g23 deletion ensitrelvir-resistant mutant. *Mbio*, pp. e00584–26, 2026.

Qisheng Pan, Stephanie Portelli, Thanh Binh Nguyen, and David B Ascher. Systematic evaluation of computational tools to predict the effects of mutations on protein-ligand binding affinity in the absence of experimental structures. *Briefings in Bioinformatics*, 27(1):bbag035, 2026.

Douglas EV Pires, Tom L Blundell, and David B Ascher. Platinum: a database of experimentally measured effects of mutations on structurally defined protein–ligand complexes. *Nucleic acids research*, 43(D1):D387–D391, 2015.

Douglas EV Pires, Tom L Blundell, and David B Ascher. mcsm-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Scientific reports*, 6(1):29575, 2016.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Big data meets quantum chemistry approximations: the δ-machine learning approach. *Journal of chemical theory and computation*, 11(5):2087–2096, 2015.

Tingting Sun, Yuting Chen, Yuhao Wen, Zefeng Zhu, and Minghui Li. Prempli: a machine learning model for predicting the effects of missense mutations on protein-ligand interactions. *Communications biology*, 4(1):1311, 2021.

Michael Tynes, Wenhao Gao, Daniel J Burrill, Enrique R Batista, Danny Perez, Ping Yang, and Nicholas Lubbers. Pairwise difference regression: a machine learning meta-algorithm for improved prediction and uncertainty quantification in chemical search. *Journal of chemical information and modeling*, 61(8):3846–3857, 2021.

Ming Hsiu Wu, Ziqian Xie, Shuiwang Ji, and Degui Zhi. Towards precision protein-ligand affinity prediction benchmark: A complete and modification-aware davis dataset. *Advances in Neural Information Processing Systems*, 38, 2025.

Yunzhuo Zhou, YooChan Myung, Alex GC de Sá, and David B Ascher. Transfer learning across molecular graphs for predicting protein-ligand affinities and their changes upon mutations. *bioRxiv*, pp. 2025–07, 2025.