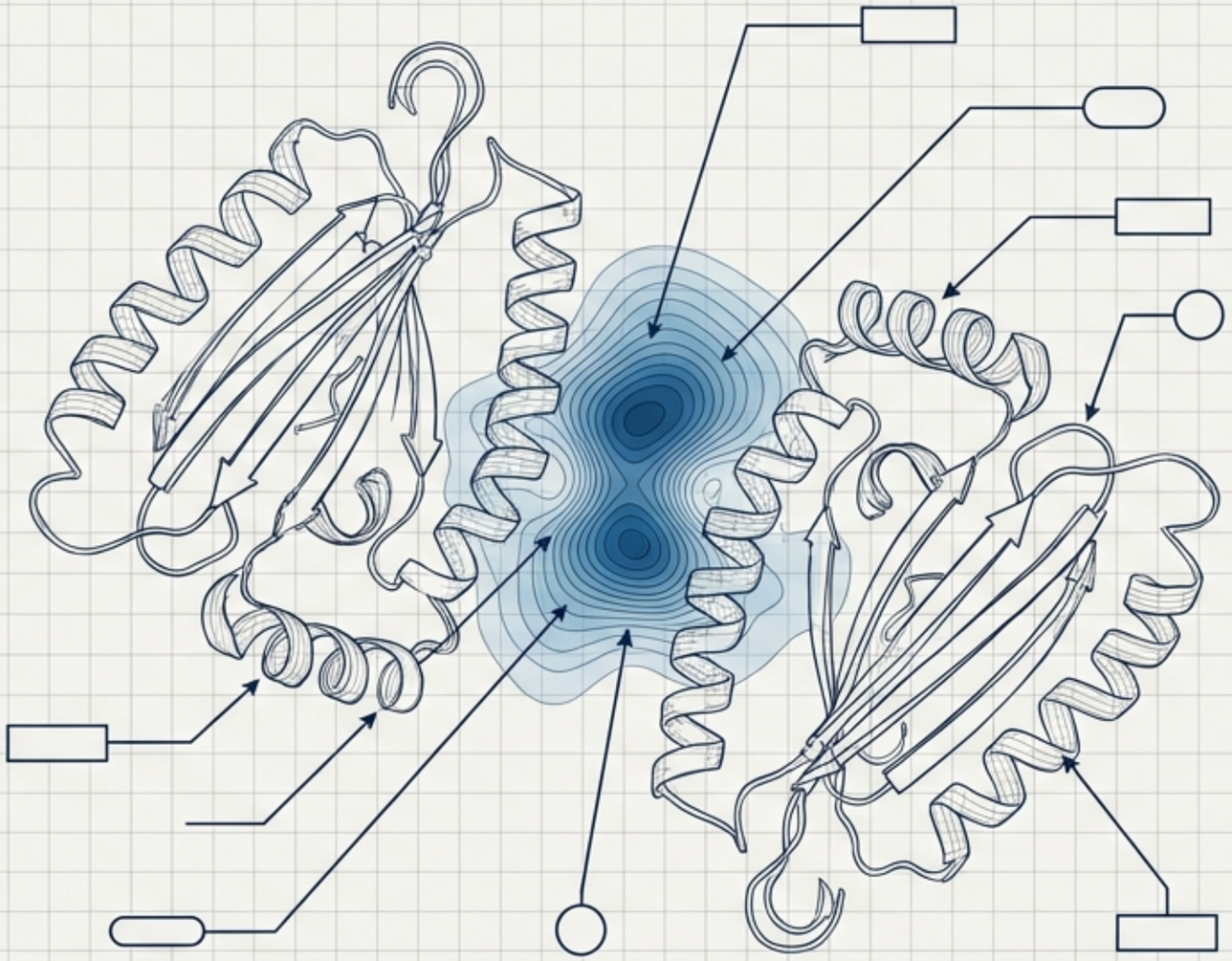
Proffinity:

Machine Learning for Affinity Prioritization of Designed Proteins

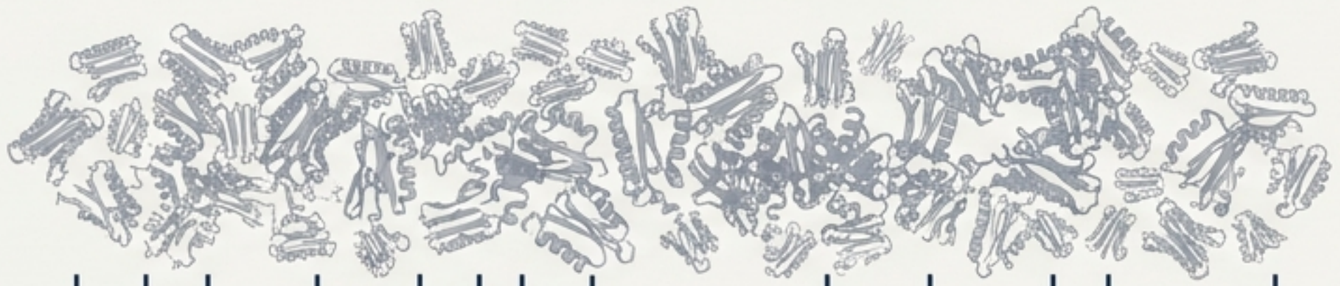
A robust, interpretable workflow to filter *de novo* generated candidates into highly specific, experimentally validated molecular outputs.



The Generative Haystack

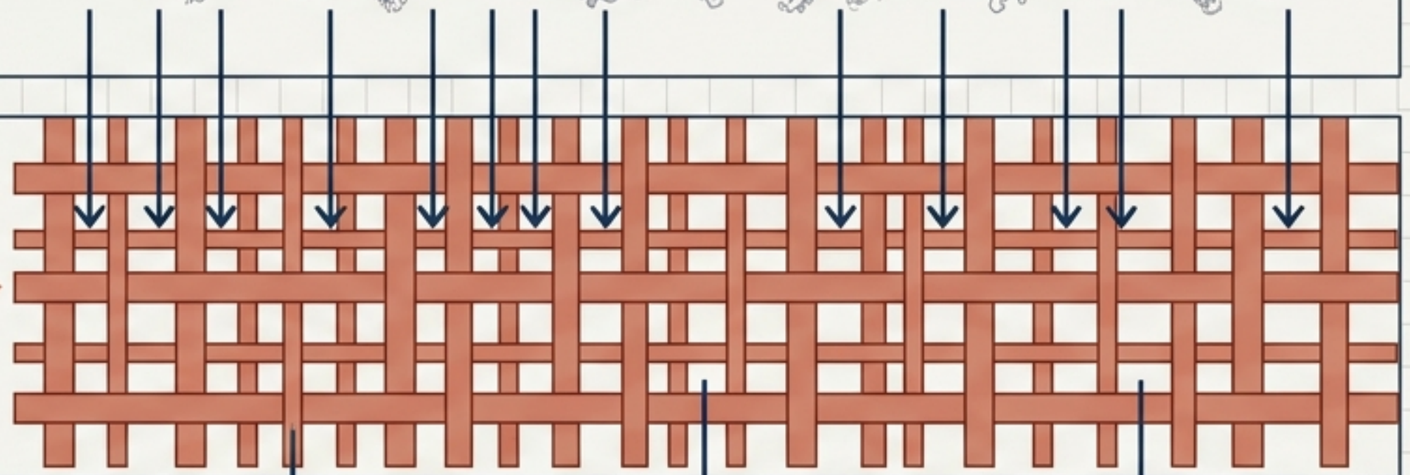
Generative AI Output

Generative models (ProteinMPNN, AlphaFold) synthesize thousands of de novo structures.



The Selection Gap

Selecting the highest-affinity binders for wet-lab validation remains a costly, high-failure bottle-neck. Unfiltered large-scale screening yields only a 1-10% experimental success rate.



The Goal: High-Affinity Binders for Wet-Lab Testing

Identified candidates with optimized binding properties ready for experimental validation.

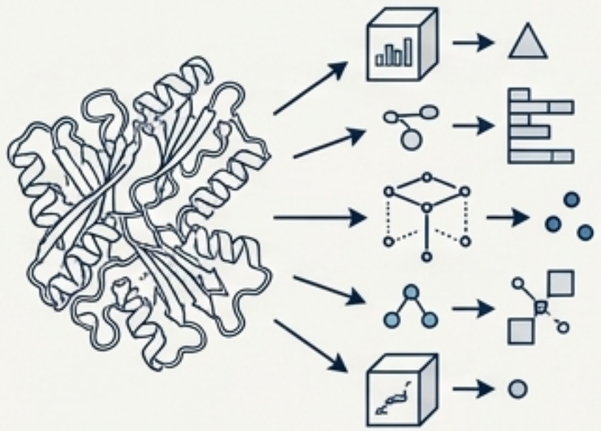
The Scoring Gap in Candidate Prioritization

Generative Metrics (pLDDT, pTM)	Legacy Empirical (FoldX, Rosetta)	Geometric Predictors (PRODIGY)	Proffinity (The Solution)
Focus: Measures structural confidence.	**Focus:** Computes pairwise atomic interactions.	**Focus:** Analyzes interfacial contact geometry.	**Focus:** Optimized for de novo design sensitivity.
Limitation: Does not predict binding affinity.	**Limitation:** Highly sensitive to structural prep artifacts; fails on multi-point mutations.	**Limitation:** Lacks residue-level biochemical nuance.	**Advantage:** Robust to structural artifacts, trained for multi-point mutations, delivers residue-level interpretability.

The Proffinity Pipeline

Step 1: Featurization

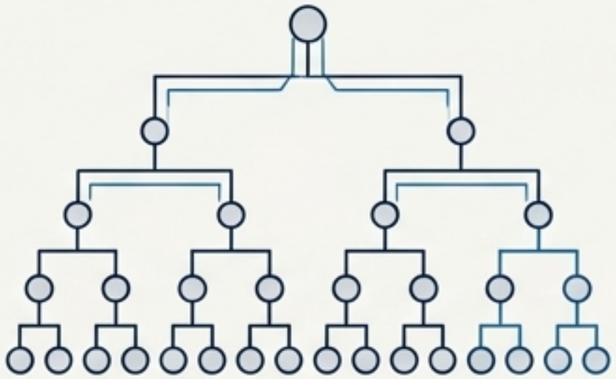
Energy Extraction



Extracts 47 distinct energy scores using knowledge-based statistical potentials.

Step 2: Machine Learning Engine

Affinity Prediction

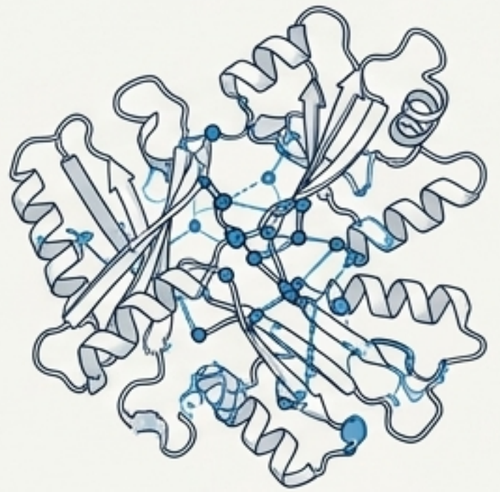


Extra Tree Regressor

Analyzes extracted features using an explainable ensemble model tuned for *de novo* mutations.

Step 3: Visualization

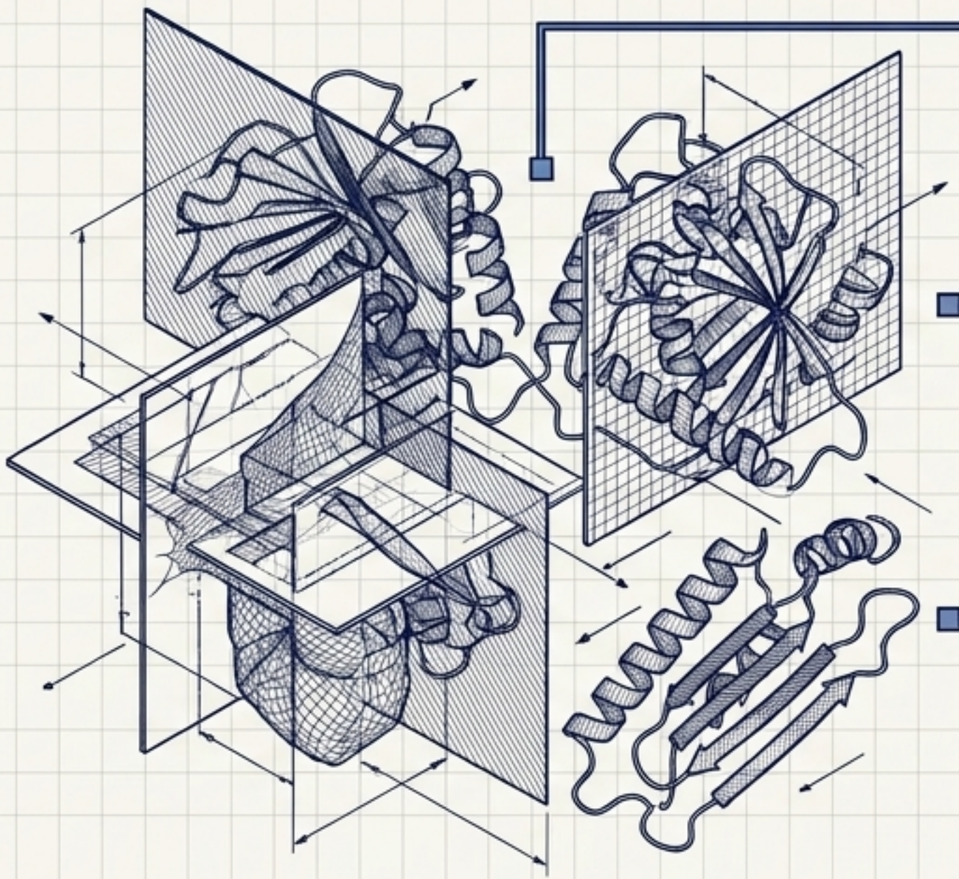
Residue Interpretability



Extrapolates detailed residue-level interactions to explain specific binding mechanics.

A one-stop computational workflow combining knowledge-based statistical potentials with explainable machine learning to accurately prioritize high-affinity binders.

Step 1: Structural Featurization



Interface Non-Bonded

Evaluated via Ca-Ca and Cb-Cb distances (<7Å). Captures the crucial non-covalent contacts across the protein-protein interface.

Fold Non-Bonded

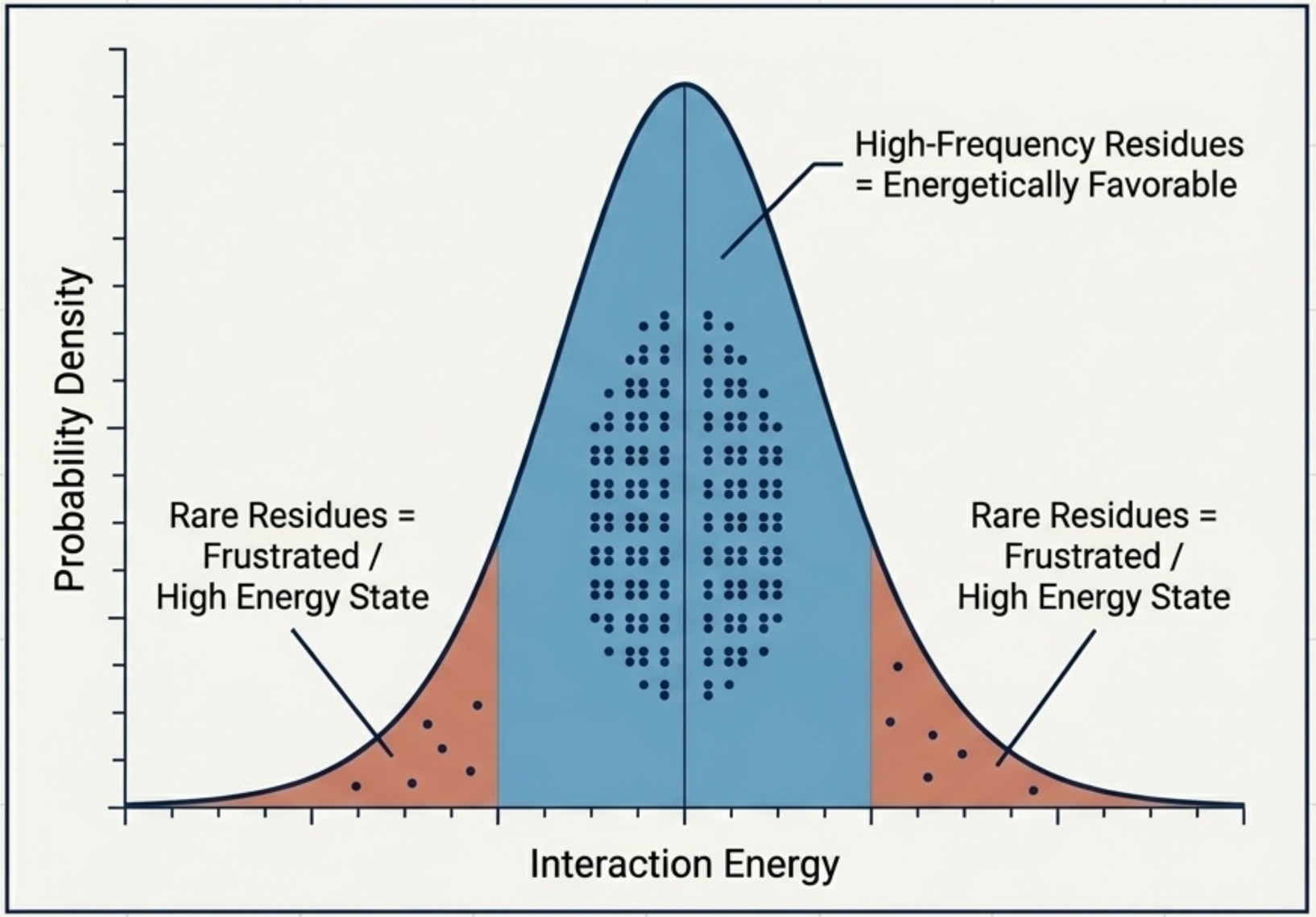
Captures intra-residue structural stability within the protein fold itself.

Fold Bonded

Captures secondary conformational states based on specific ψ-ϕ angles in existing sequences.

Proffinity extracts 47 distinct energy scores using 11 **statistical potentials** derived from known PDB structures, capturing a wide spectrum of **non-covalent and covalent interactions**.

The Math of Interaction: Statistical Contact Potential

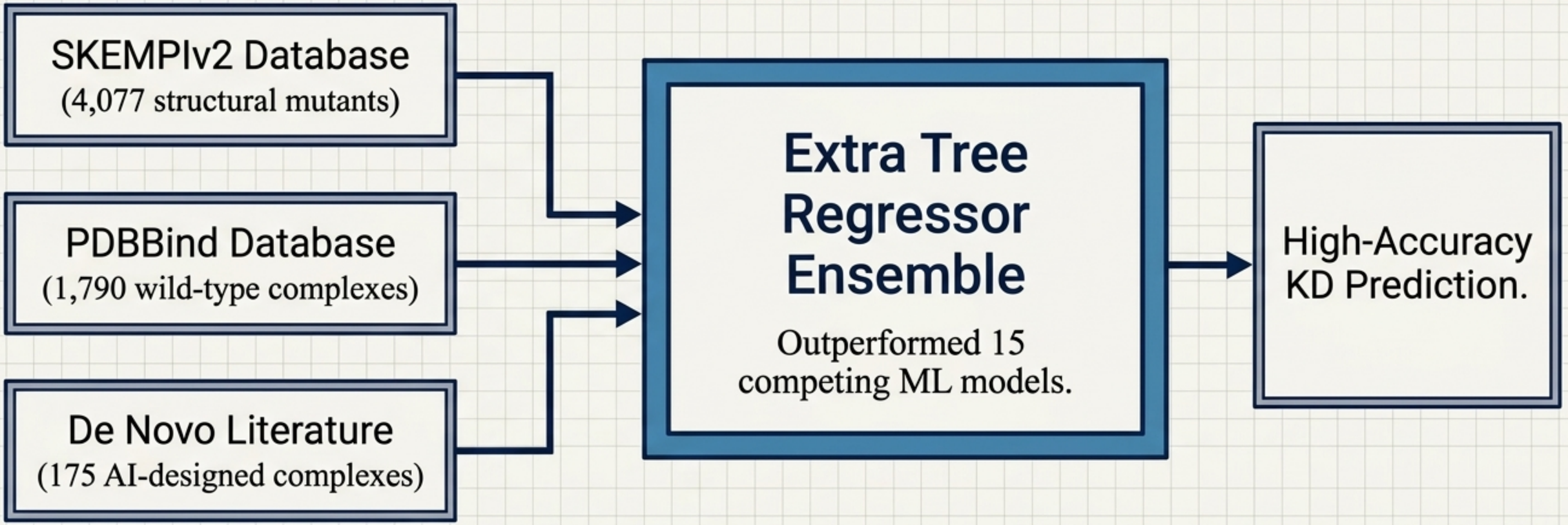


How frequency translates to affinity.

Statistical contact potential utilizes a Boltzmann distribution to translate the observed frequency of residue pairs in nature nature into an actionable binding affinity score.

Common pairs stabilize the complex; rare pairs introduce structural frustration.

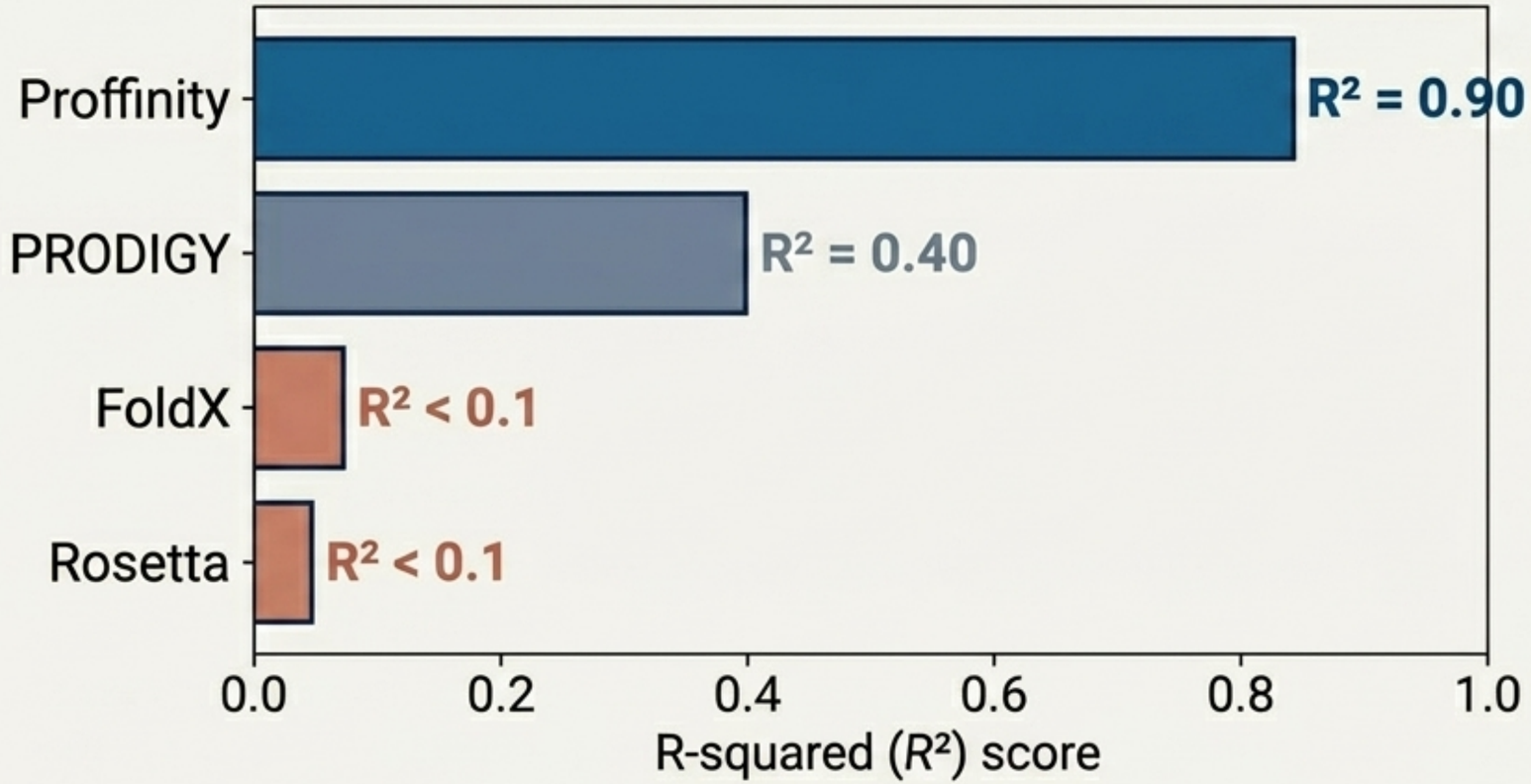
Step 2: The Machine Learning Engine



Uniquely trained for multi-point mutation sensitivity, the model accurately detects nuanced binding behaviors that escape standard linear algorithms, specifically tailoring it to the realities of de novo protein design.

Benchmark Dominance

Evaluating performance against the SKEMPIv2 holdout set.



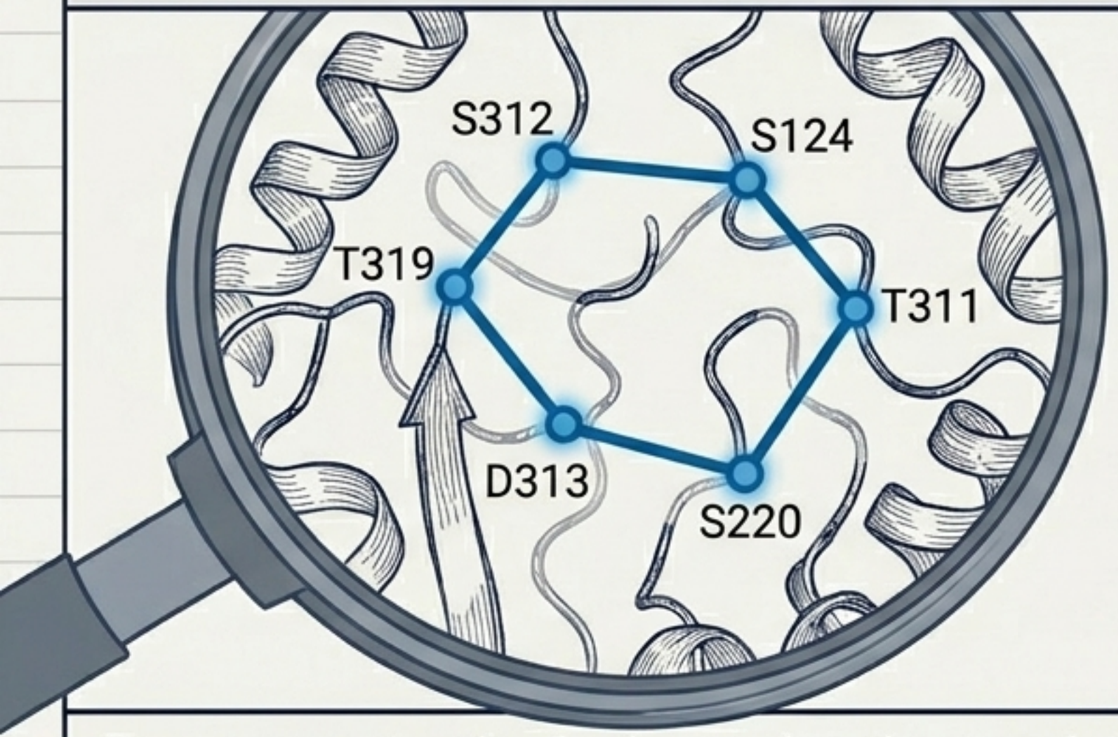
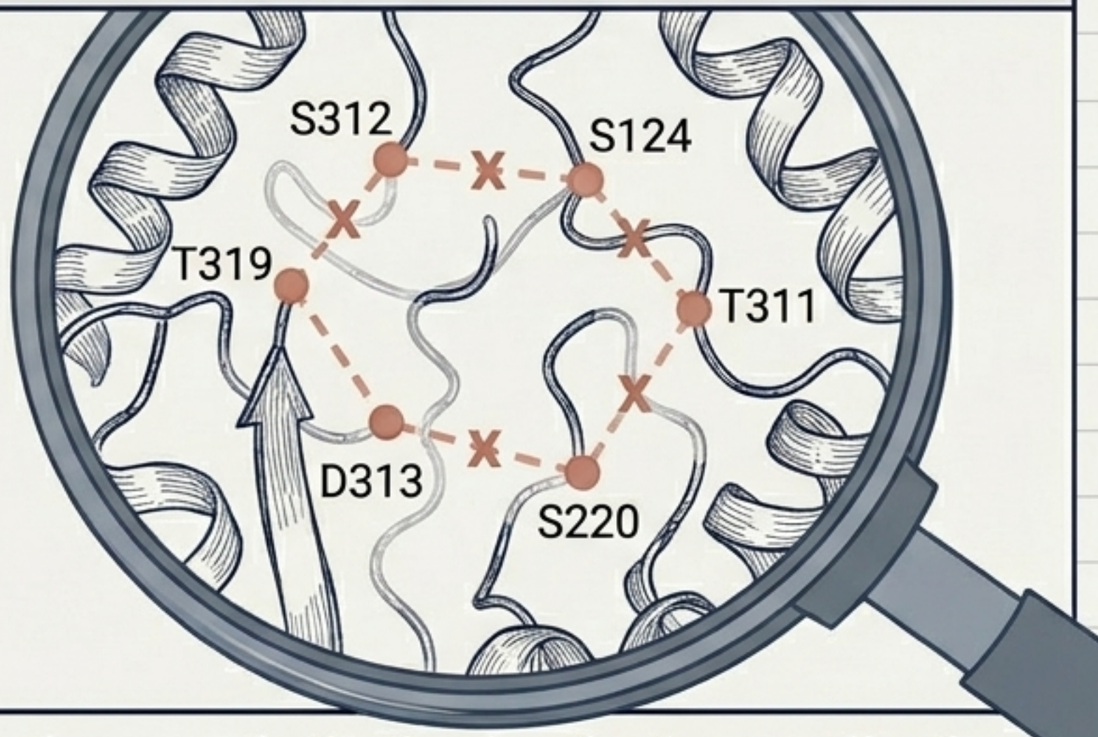
Proffinity vastly outperforms commonly used PPI affinity predictors. It maintains robust accuracy even on homology-modeled structures where legacy geometric and empirical tools (FoldX, Rosetta) fail due to structural prep sensitivity.

Step 3: Residue-Level Interpretability

$$[\text{Raw Energy Score}] \times [\text{ML Feature Importance Weight}] = \frac{WI_S}{W_{RS}}$$

WRS (Weighted Residue Score)	WIS (Weighted Interaction Score)
Aggregates bonded and non-bonded interactions to score the overall stability of an individual residue within the complex.	Maps highly connected residue-residue interaction networks to explicitly explain the structural “why” behind a sequence’s binding affinity.

Interpretability in Action: Eglinc Mutants

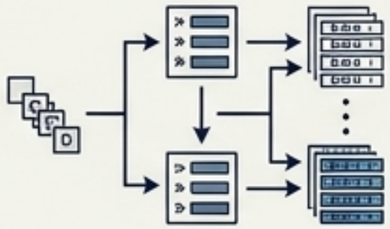
L312S Mutant	L312I Mutant
WRS: 1.0 | K_D: 15.5 µM	WRS: 0.4 | K_D: 57.6 µM
	
Forms an extensive hydrogen bond network (S312-S124, S124-T311, T311-S220, S220-D313, D313-T319), uniquely introducing the S220-D313 interaction.	Homologous substitution breaks the coordination, resulting in a fractured interaction network and significantly weaker binding.

Proffinity mathematically isolates the specific structural mechanisms driving affinity.

The Road Test: Designing Rsp5-UbV Binders

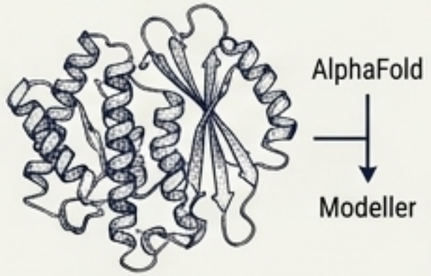
Target: The yeast Rsp5 HECT-type E3 ligase.
Goal: Design high-affinity Ubiquitin Variants (UbVs) capable of allosteric activation.

1 Generate Candidates



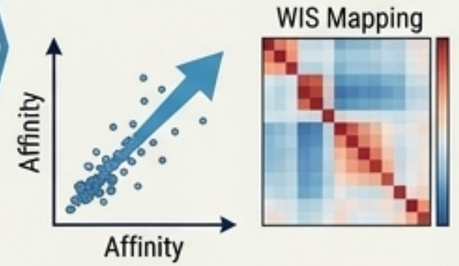
1,000 *de novo* sequences proposed via ProteinMPNN.

2 Predict Structures



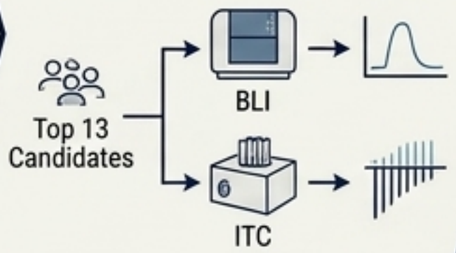
Complex geometries modeled via AlphaFold & Modeller.

3 Proffinity Filtering



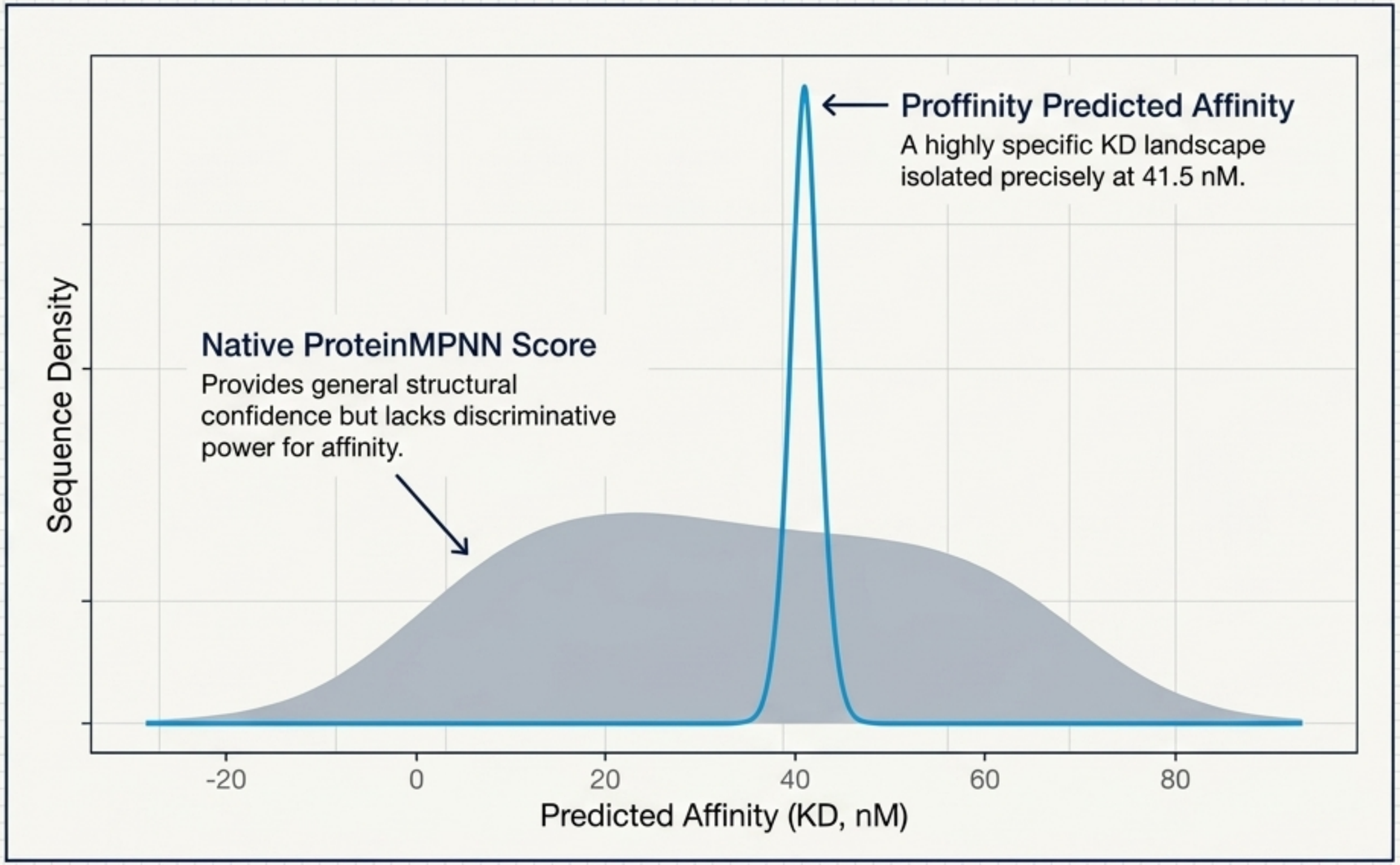
Library prioritized via precise KD affinity prediction and WIS mapping.

4 *In Vitro* Validation



Top 13 selected candidates proceed to physical testing via BLI/ITC.

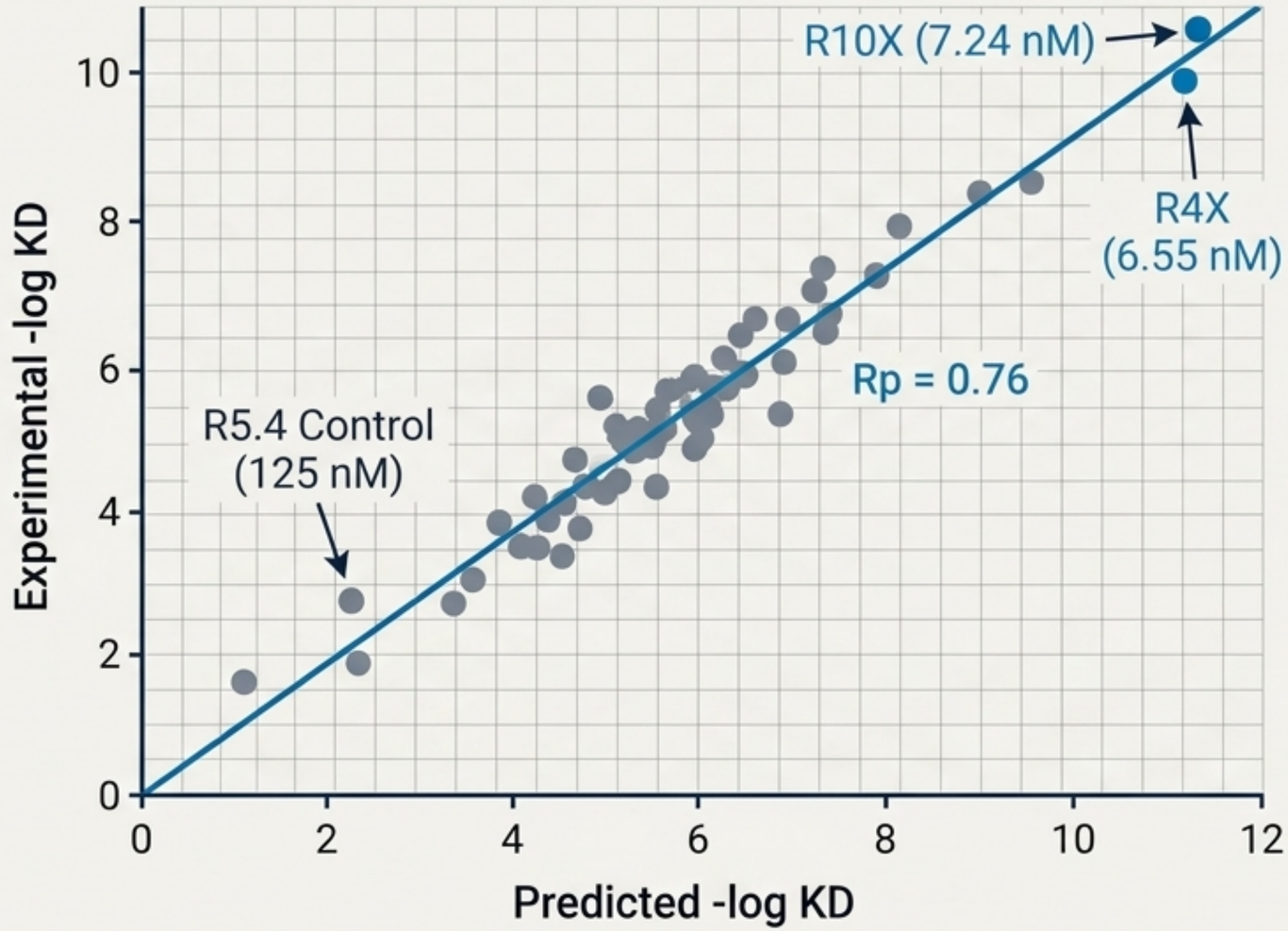
Narrowing the Generative Haystack



Transforming a blind guess into a targeted selection.

While generative sequence metrics confirm that a candidate will fold, Proffinity maps the exact predicted binding affinity, *allowing researchers to surgically isolate the few true interactors from thousands of false positives.*

Experimental Validation



Unprecedented Accuracy

11 of 13 candidates selected by Proffinity demonstrated exceptionally tight binding (< 1 μM) in wet-lab validation.

Massive Affinity Gains

The top combinatorially designed interactors easily beat the phage-display control, exhibiting **up to a 100-fold enhancement over previous unguided designs.**

Confirming Target Specificity

A Site Overlap Confirmed

Epitope Binning via BLI proves exact exosite targeting. R5.4-saturated Rsp5 rejects new UbVs.

B E2 Independence Maintained

Proving functional viability. The E2 site remains fully accessible for catalytic function even when the exosite is bound.

Proffinity guarantees structurally precise binders that maintain the complex functional geometry required for allosteric activation.

The Filtered Paradigm

Generative Only

- High wet-lab costs
- Trial-and-error validation
- Blind structural assumptions

1–10% Success Rate

Generative + ML Filtering

- Fast, targeted validation
- Interpretable mechanisms
- Residue-level precision

~50% Success Rate

</> Open-Source and Ready for Implementation.
github.com/yuchen-lo/Proffinity