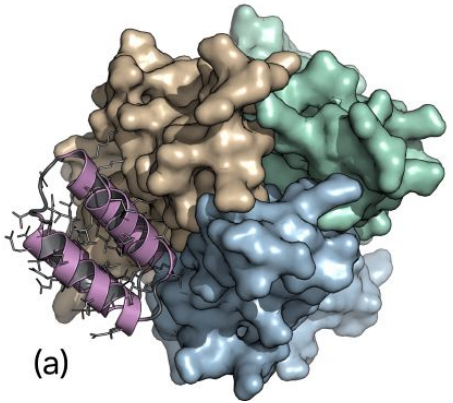
Proteína-Complexa

Scaling Atomistic Protein Binder Design with Generative Pretraining and Test-Time Compute

Kieran Didi et al.
ICLR 2026



Erik Dyer and Rishi Kannan
COS 597Y

Didi et al. (2026), ICLR 2026. Figure 2a, p. 3

Agenda

1. Motivation and scientific context
2. Background and related work
3. Problem formulation and research question
4. Technical overview of the proposed method
5. Experimental setup and evaluation
6. Strengths, limitations, and open questions
7. Discussion: What did this paper teach us?
   What are promising future directions?

Motivation and scientific context

1 **Motivation and scientific context**

2 Background and related work

3 Problem formulation and research question

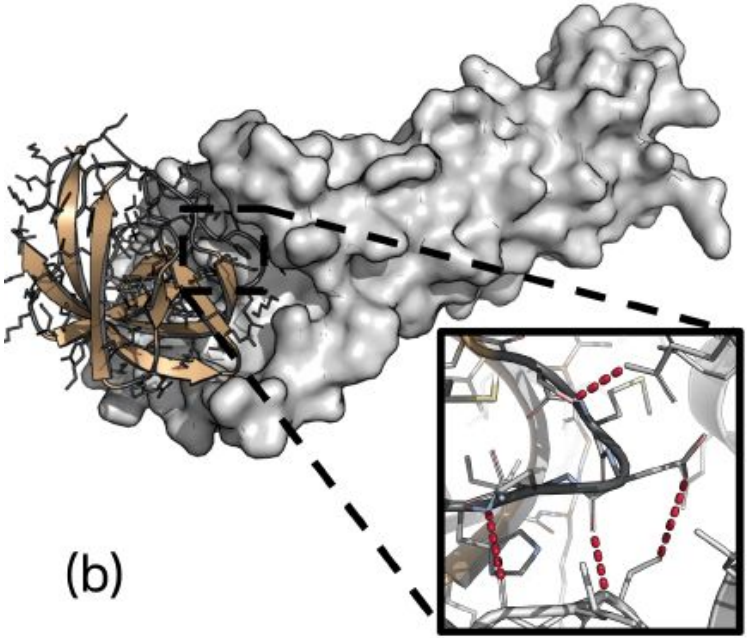
4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

Why Design Protein Binders?



Molecular recognition

The binder must fold and contact the target in the intended pose.

Atomic geometry and chemistry

Side-chain packing and hydrogen bonds shape the interface.
Drug discovery uses these contacts to block or modulate target proteins.

Claudin-1 target and designed binder. Red marks show interface hydrogen bonds.

Figure 2b, p. 3. Motivation: §1

Background and related work

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

Generative design and hallucination

Design approach	Generative models	Hallucination methods
Examples	RFDiffusion, APM	BindCraft, BoltzDesign
Candidate source	Sample a learned generator	Optimize a binder sequence
Role of predictor	Evaluate designs	Guide sequence optimization

Complexa combines a learned generator with optimization guided by a structure-prediction reward.

Background and related work: §2, p. 3

Problem formulation and research question

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

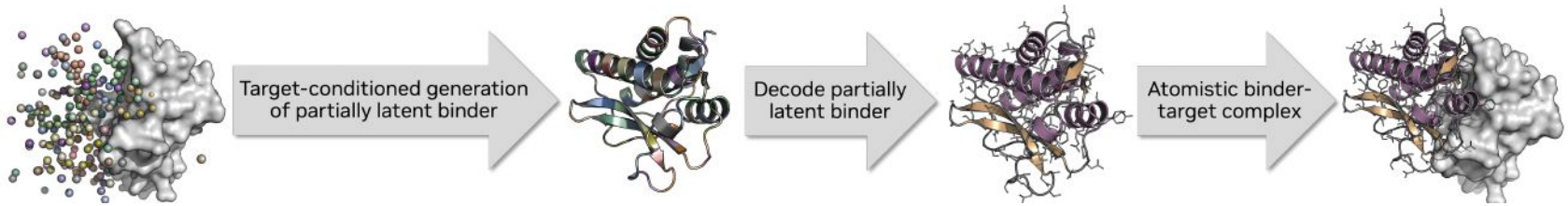
4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

Target-Conditioned Generation



Given

Target structure, hotspot residues, and binder length

Generate

Binder sequence and all-atom structure, generated jointly

Complexa combines generative proposals with reward-guided optimization of their trajectories.

Figure 1, top, p. 2. Target and hotspot conditioning: §3.2

Generative Pretraining and Test-Time Compute

Generative Pretraining

Can domain-domain interactions supply useful binder-target training data?

Inference-time optimization

When does trajectory search improve on independent sampling?

In-silico success

What do predictor-based metrics establish about binding?

Technical overview of the proposed method

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

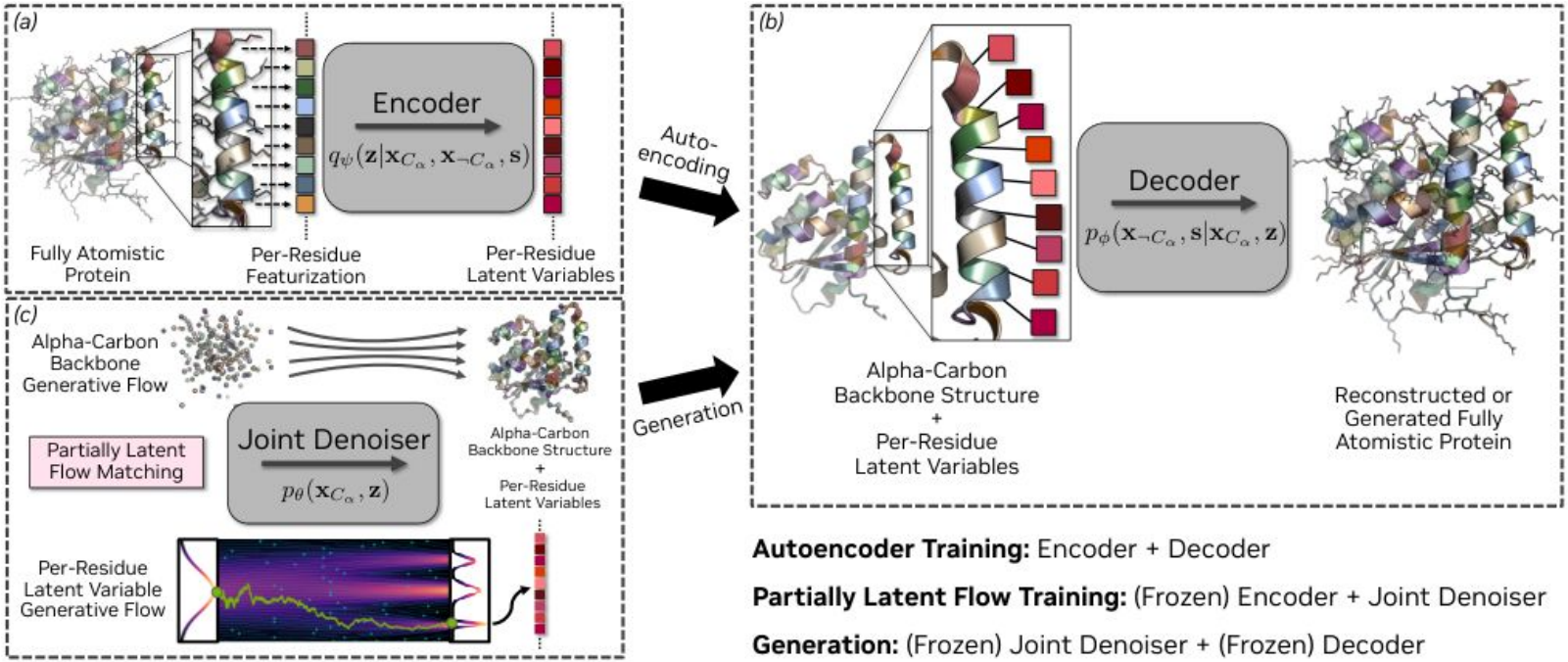
4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

Predecessor: La-Proteína Model Architecture



Shared core: transformers with pair-biased attention.
Generation uses the joint denoiser, then the decoder.

La-Proteína Figure 1 and §3.3 (arXiv:2507.09466v2). Core inherited by Complexa.

Partially Latent Representation

Explicit geometry

Cα coordinates: L × 3

Learned atomistic content

Latent variables z: L × 8
Sequence and non-Cα atoms

$$p_\phi(\mathbf{x}^{\neg C_\alpha}, \mathbf{s} | \mathbf{x}^{C_\alpha}, \mathbf{z}) = p_\phi(\mathbf{s} | \mathbf{x}^{C_\alpha}, \mathbf{z}) p_\phi(\mathbf{x}^{\neg C_\alpha} | \mathbf{x}^{C_\alpha}, \mathbf{z}), \quad (5)$$

Sequence probabilities

Categorical distribution

Non-Cα coordinates

Gaussian coordinate means

Appendix B.1–B.2, pp. 19–20. Exact Equation (5)

Flow Matching

$$\boldsymbol{x}_t = t\,\boldsymbol{x}_1 + (1-t)\,\boldsymbol{x}_0$$

t = 0: noise
t = 1: protein

Learn the velocity along a path from noise to a training protein.

$$\min_\theta \mathbb{E}[\boxed{||\mathbf{v}_\theta^x(\mathbf{x}_{t_x}^{C_\alpha}, \mathbf{z}_{t_z}, t_x, t_z) - (\mathbf{x}^{C_\alpha} - \mathbf{x}_0^{C_\alpha})||^2} + \boxed{||\mathbf{v}_\theta^z(\mathbf{x}_{t_x}^{C_\alpha}, \mathbf{z}_{t_z}, t_x, t_z) - (\mathbf{z} - \mathbf{z}_0)||^2}], \quad (7)$$

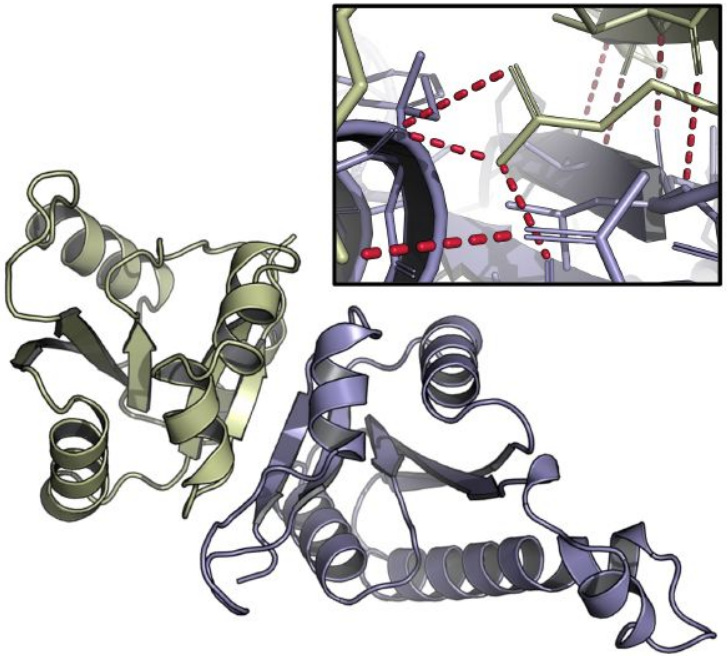
Cα velocity error

Latent velocity error

Why use different schedules for coordinates and latents?

Rectified flow: §2, p. 3. Exact Equation (7), Appendix B.3, p. 20

Teddymer: Interacting Protein Domains



Experimental structures

45,856 filtered PDB multimers provide experimentally resolved complex structures.

Synthetic interaction data

510,454 filtered Teddymer representatives (~11× the PDB set), from contacting domains.

These are training examples rather than experimentally tested designed binders

Teddymer: Figure 3, §3.1 and Appendix C. Filtered PDB count: Appendix D.1.

Teddymer Construction and Filtering

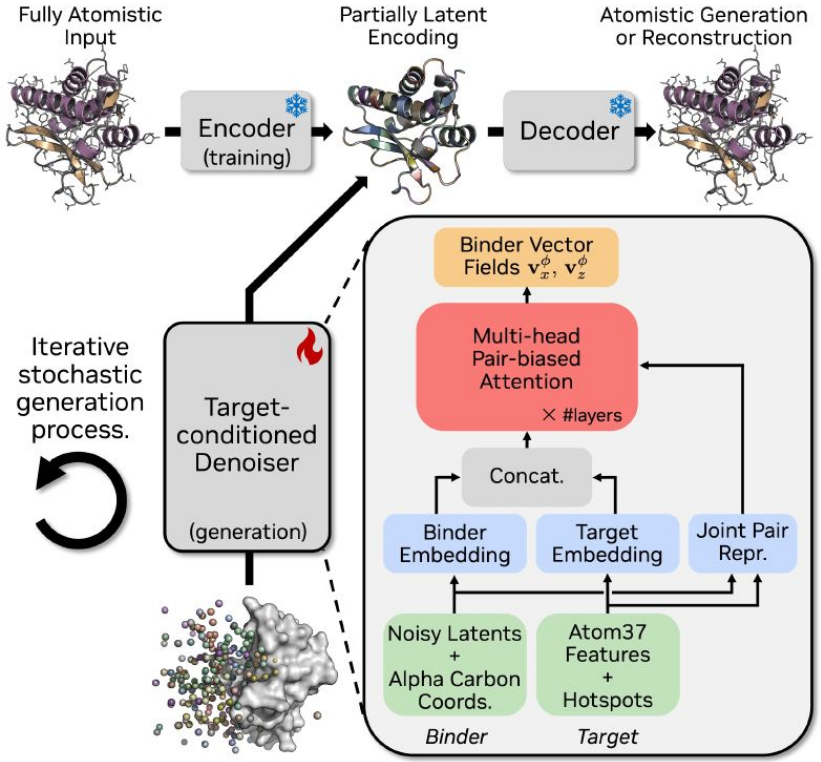
Stage	Count
AFDB50 structures with TED domains	47,180,623
Spatially adjacent domain pairs	123,606,001
Pairs with C.A.T. annotations	10,089,503
Dimer structural clusters	3,556,223
Filtered representatives used for training	510,454

Adjacent domains become dimers, which are grouped by structural similarity.

Quality filtering keeps representative interfaces for training.

§3.1 and Appendix C, p. 21. Figure 4, p. 4

Complexa: Target-Conditioned Denoiser



Clean target and noisy binder

Target atoms and hotspots condition the binder trajectory.

Pair-biased transformer

Distances bias binder–target attention; outputs are Cα and latent velocities.

The inherited encoder and decoder stay frozen - target conditioning trains the denoiser.

Exact Figure 5, p. 5. Architecture: §3.2 and Appendix G.1–G.2

Translation Noise

$$\boldsymbol{x}_{t_x}^{C\alpha} = t_x\, \boldsymbol{x}^{C\alpha} + (1 - t_x)\left(\boldsymbol{x}_0^{C\alpha} + \vec{d}\,\mathbf{1}_N\right)$$

Applied to every noisy Cα position
Standard deviation: 0.2 nm per coordinate

Shared translation d

Without the shift

Zero-centered noise leaks the clean binder's center of mass through t.

Why add it?

The random shift forces the model to learn placement relative to the fixed target.

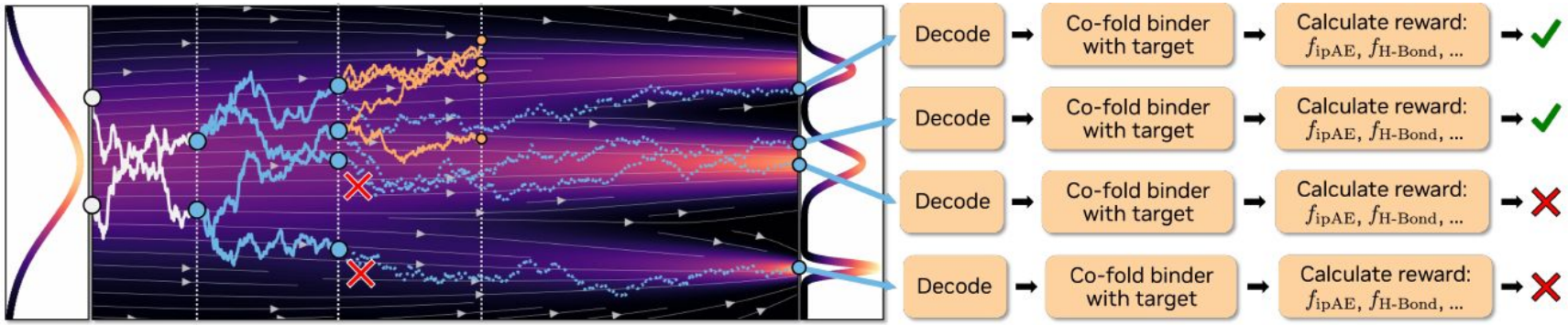
Translation-noise interpolant from Equation (1), pp. 5–6. Training: Appendix G.3, p. 28

Inference-Time Optimization Algorithms

Algorithm	How it allocates compute
Best-of-N	Independent complete samples
Beam search	Top-scoring partial trajectories
Feynman–Kac steering	Reward-weighted resampling
MCTS	Tree search with exploration
Generate and Hallucinate	Generated sequence, then BindCraft refinement

Search algorithms: §3.4, pp. 7–8 and Appendix H

Beam Search: A Worked Example



1. Branch partial states

Continue several noisy Cα + latent states; finish rollouts to estimate reward.

2. Score and retain

Decode, refold, and score each rollout. Keep the best partial states and repeat.

Figure 1, bottom, p. 2. Beam search: §3.4, p. 7 and Appendix H.2

Experimental setup and evaluation

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

In-Silico Success Metrics

Metric	What it measures	Protein threshold
Interface pAE	Confidence in relative chain placement	< 7 Å
Complex pLDDT	Confidence in local structure	> 90 on a 0–100 scale
Binder scRMSD	Agreement between design and refolding	< 1.5 Å

These criteria screen refolded designs for the computational benchmarks.

What would an independent binding assay add?

Protein criteria: §3.3, p. 6 and Appendix F, pp. 24–25

Base Generative Model Benchmark

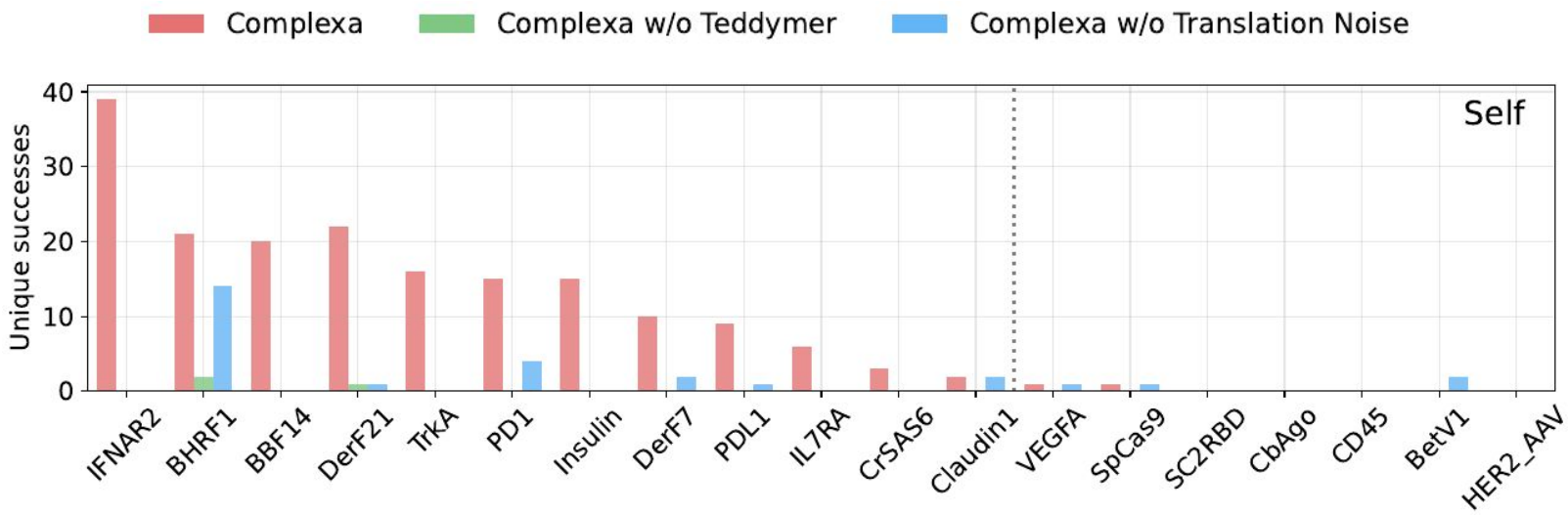
200 designs per target - 19 protein targets

Method	Own sequence[1]	MPNN redesign[1]	Generation time[2]
RFDiffusion	—	4.68	70.8 s
Protpardelle-1c	—	0.73	8.13 s
APM	0.31	3.15	73.1 s
Complexa	9.10	14.4	15.6 s

[1] Mean unique successes across targets [2] Seconds per sample; excludes evaluation

Table 2, p. 8. Fixed sample budget: 200 designs per target, 19 protein targets

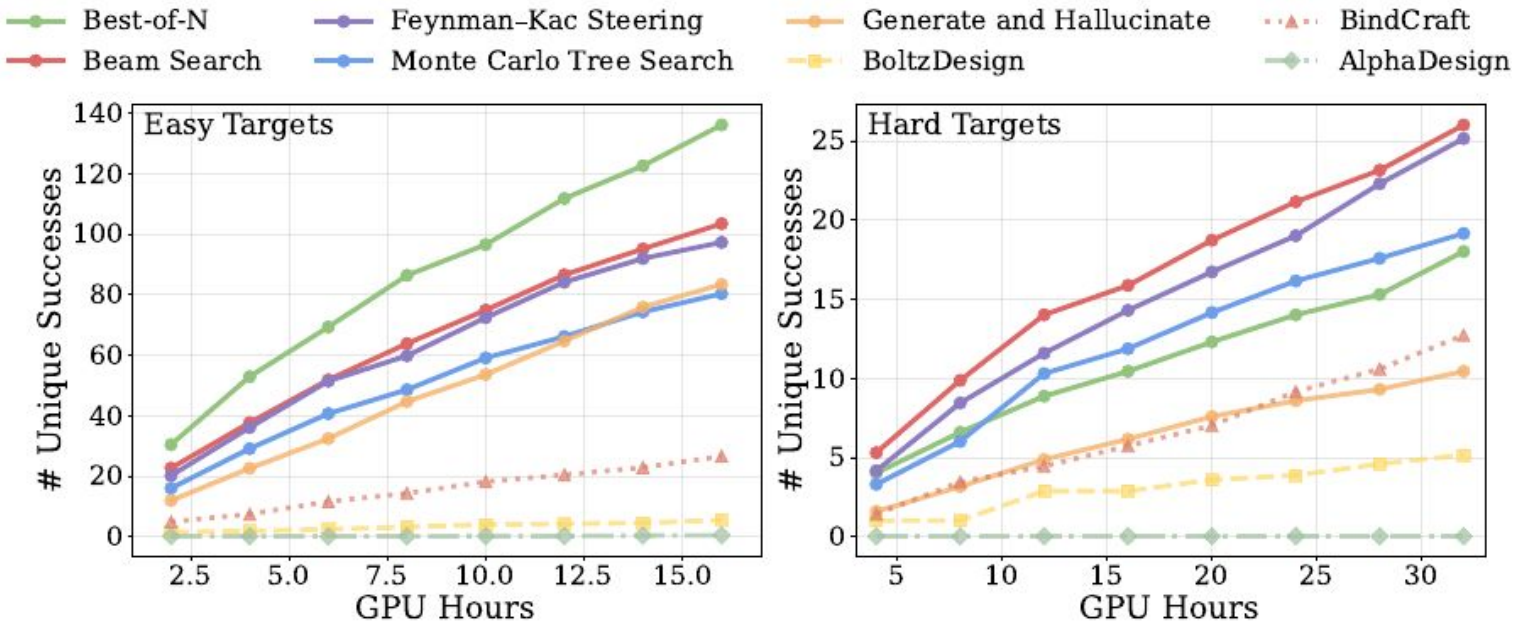
Ablations on Data and Translation Noise



Each bar group is a target; height counts unique successes from its own sequences.
Colors compare the full model with training that omits Teddymer or translation noise.
Both ablations reduce success; easy targets are left of the dashed line, hard targets right.

Figure 13, Self panel, p. 35. Table 6, p. 34

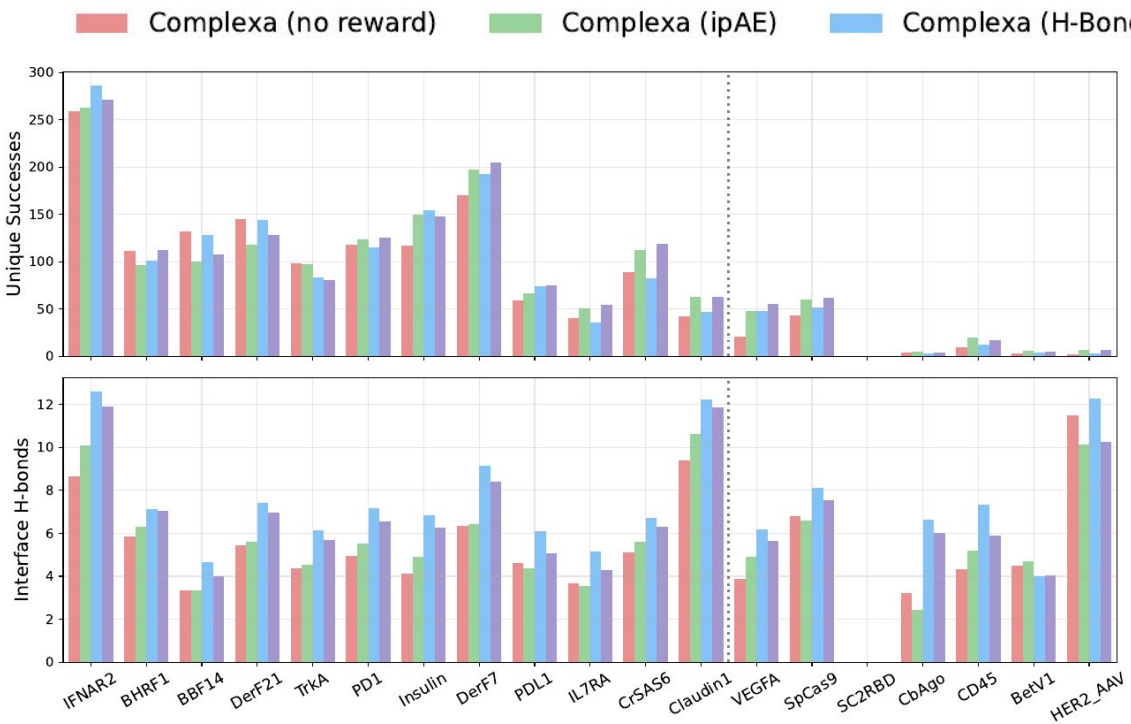
Impact of Search on Easy vs Hard Targets



12 easy / 7 hard targets; mean unique successes versus total inference GPU hours. Best-of-N leads on easy targets; guided search improves the hard-target mean.

Exact Figure 7, p. 9. Definitions: Appendix E. Time accounting: Figure 21–22 captions

Interface Hydrogen-Bond Optimization



- Each x-axis label is a different protein target (19 total).
- Top: unique successes.
- Bottom: mean interface H-bonds among successful refolded designs.
- Colors: no reward, ipAE, H-bonds, or both.
- The reward changes interface properties; binding still needs an assay.

Figure 24, p. 43. Hydrogen bond counts on refolded successful designs. Appendix I.8

Strengths, limitations, and open questions

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

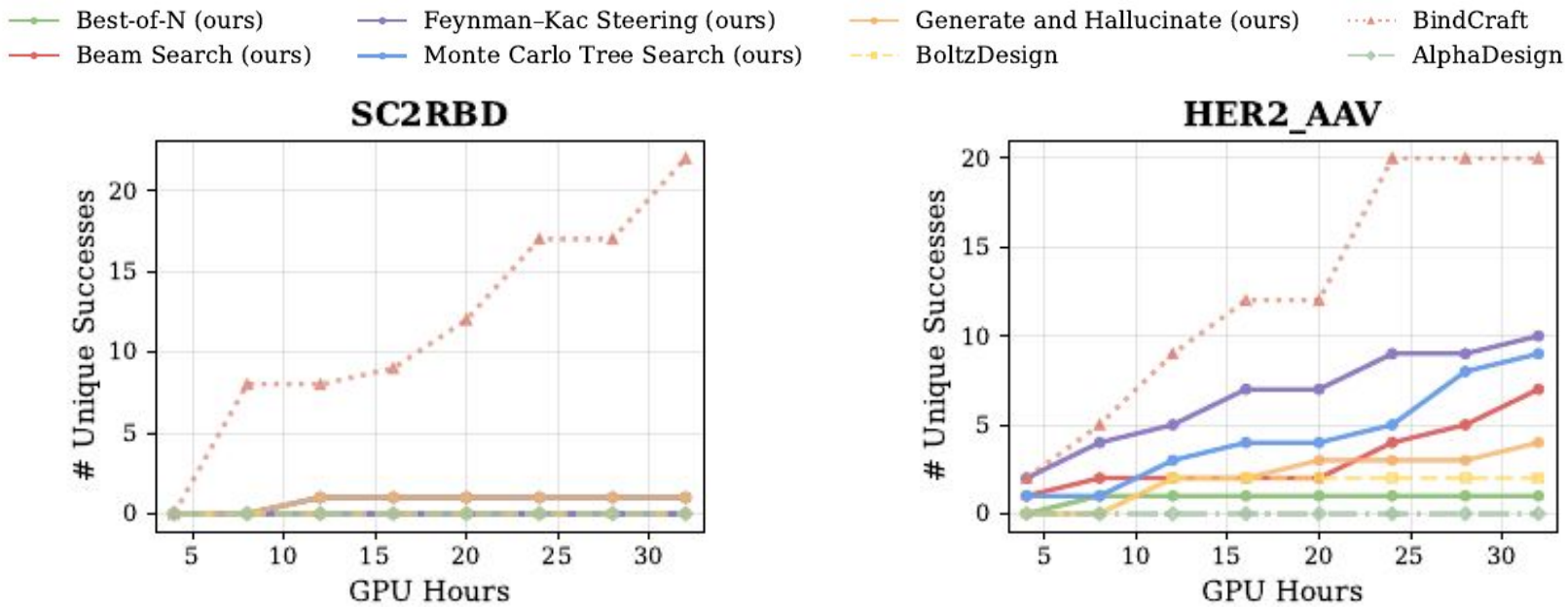
4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us?
What are promising future directions?

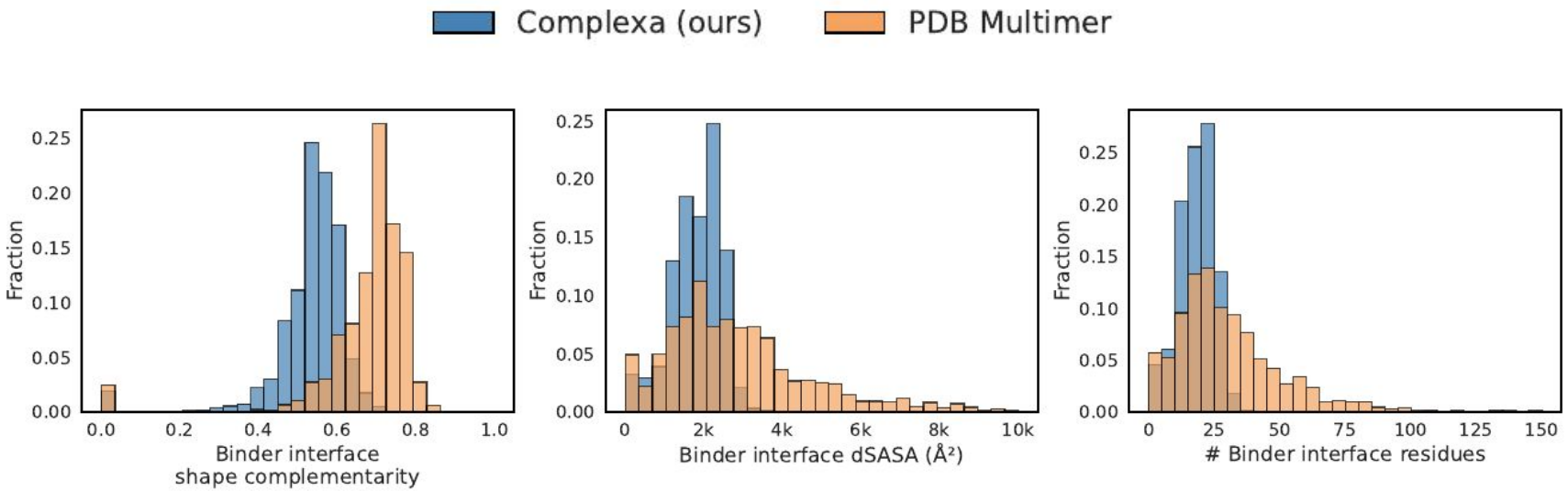
Per-Target Search Results



Each panel is one target: SC2RBD (left), HER2_AAV (right).
BindCraft performs better here; the hard-target average hides these exceptions.

Figure 22, selected SC2RBD and HER2_AAV panels, p. 42. Interpretation: Appendix I.6, p. 41

Biophysical Interface Properties



Blue: successful Complexa designs. Orange: experimental PDB multimers.
Left → right: shape complementarity, buried interface area, interface residues.
Complexa interfaces tend to be smaller; these plots do not measure binding affinity.

Figure 29, lower three panels, p. 50. Appendix I.13, p. 49

Strengths and Limitations

Strengths	Limitations
Scalable synthetic interaction data	Domain contacts are a proxy for binding between separate proteins
Search improves hard-target averages	Gains depend on the target; no algorithm wins everywhere
Joint atom–sequence design; flexible rewards	No wet-lab validation; affinity and specificity remain untested

Compute comparisons cover inference, not pretraining cost.
Next step: matched binding assays with equal candidate-testing budgets.

Complexa: Tables 2 and 6, Figure 7, Appendix A and I.6–I.8 (arXiv:2603.27950v1).

Discussion and future directions

1 Motivation and scientific context

2 Background and related work

3 Problem formulation and research question

4 Technical overview of the proposed method

5 Experimental setup and evaluation

6 Strengths, limitations, and open questions

7 Discussion: What did this paper teach us? What are promising future directions?

What the Evidence Supports

Evidence Supports	Evidence still Needed
Teddymer improves computational results	How strongly and selectively do designs bind?
Search improves hard-target averages	Do other tests confirm the results?
Joint sequence and structure generation	Do the designs work as intended?

Synthesis of Tables 2 and 6, Figures 7 and 22, Appendix A and I.7

Discussion

1. How should we decide when to sample and when to search?
2. What would convince us that a higher score means stronger binding?
3. How could we test the limits of the generative prior?
4. What should the reward include beyond interface confidence?
5. Which experiment would you run next?

Discussion prompts based on Appendix A, I.6–I.7, and Figure 7