

evedesign: accessible biosequence design with a unified framework

Thomas A. Hopf¹, Artem Gazizov², Sergio Garcia Busto^{3,4}, Ethan Eschbach², SunJae Lee⁵, Milot Mirdita⁶, Rose Orenbuch², Khaoula Belahsen², David Ross⁷, Chris Sander², Martin Steinegger⁵, Simon d'Oelsnitz², Debora S. Marks²

¹Thomas Hopf Scientific Consulting, Bad Schönborn, Germany
²Department of Systems Biology, Harvard Medical School, Boston, USA

³Wellcome Sanger Institute, Cambridge, UK

⁴University of Cambridge, Cambridge, UK

⁵School of Biological Sciences, Seoul National University, Seoul, Republic of Korea

⁶Department of Precision Medicine, Sungkyunkwan University School of Medicine, Suwon, Republic of Korea

⁷National Institute of Standards and Technology, Biosystems and Biomaterials Division, Gaithersburg, USA

Explore & contribute!



evedesign

How is evedesign different to existing approaches?

Capability	Single-method tools / ad hoc workflows	evedesign
Model interoperability	Each tool has its own API and data format	Standardized interface: any model can be swapped or combined
Design representation	Fixed single-level (sequence or structure or embedding)	Multi-level: sequence, embedding, and 3D structure simultaneously
Conditional design	Hardcoded per tool	Declarative system specification; constraints composable at runtime
Multi-objective optimization	Requires bespoke glue code for each combination	Native: compose score/generate/transform operations across models
Supervised + unsupervised integration	Manual, ad hoc	First-class support: attach experimental labels directly to instances
Lab-in-the-loop iteration	Not supported	Designed for it: declarative, serializable pipelines; halt, replay, and update with new data
User interface	None, or tool-specific	End-to-end: target sequence to orderable DNA, model-agnostic
Privacy & self-hosting	Varies; often unavailable	<i>evedesign_server</i> can be hosted on private infrastructure
Access & reproducibility	Proprietary or fragmented	Open-source, FAIR-compliant, self-hostable

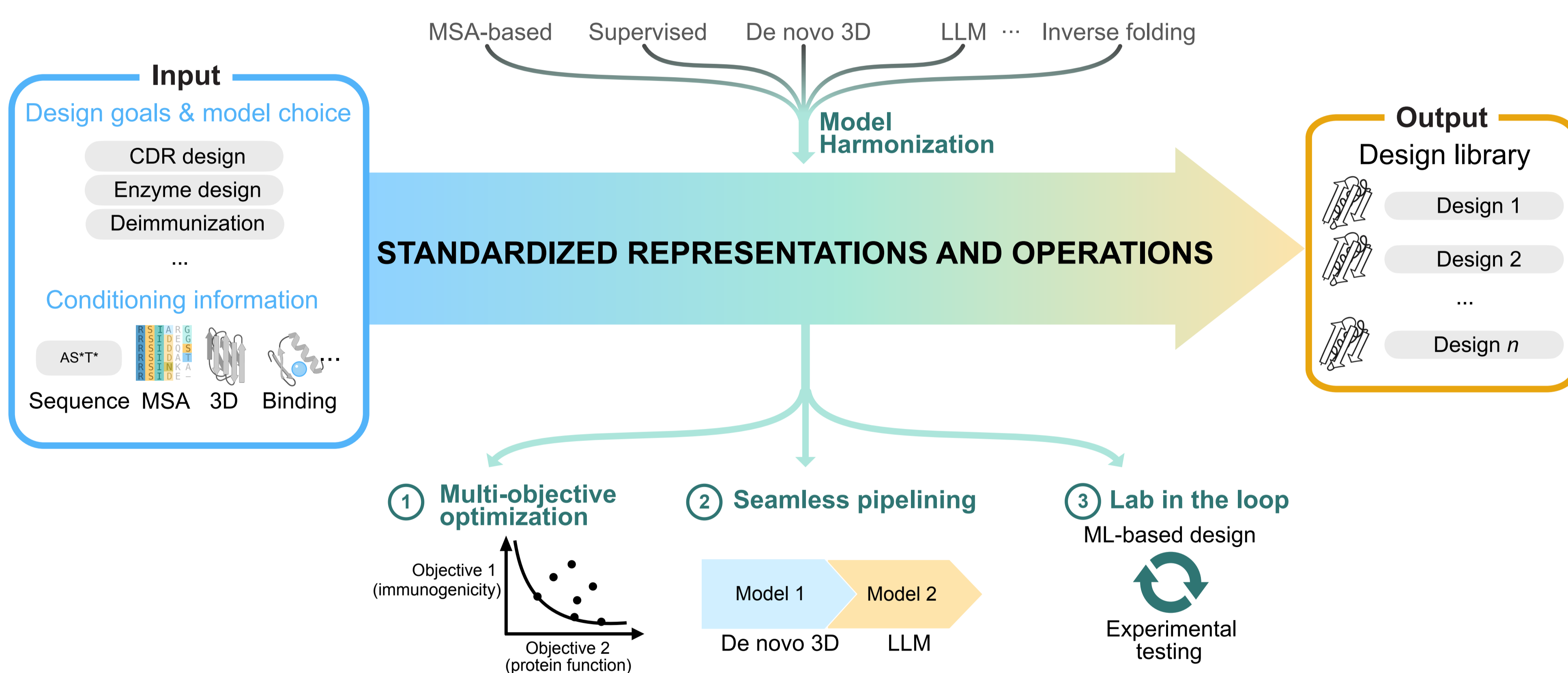
The problem: ML models for biodesign are disjointed

There is a multitude of ML models available for protein design, but they are **fragmented and task-specific**, and rarely support lab-in-the-loop workflows.

Building **pipelines** for real-world protein engineering tasks **requires bespoke glue code**, and remain inaccessible to non-experts.

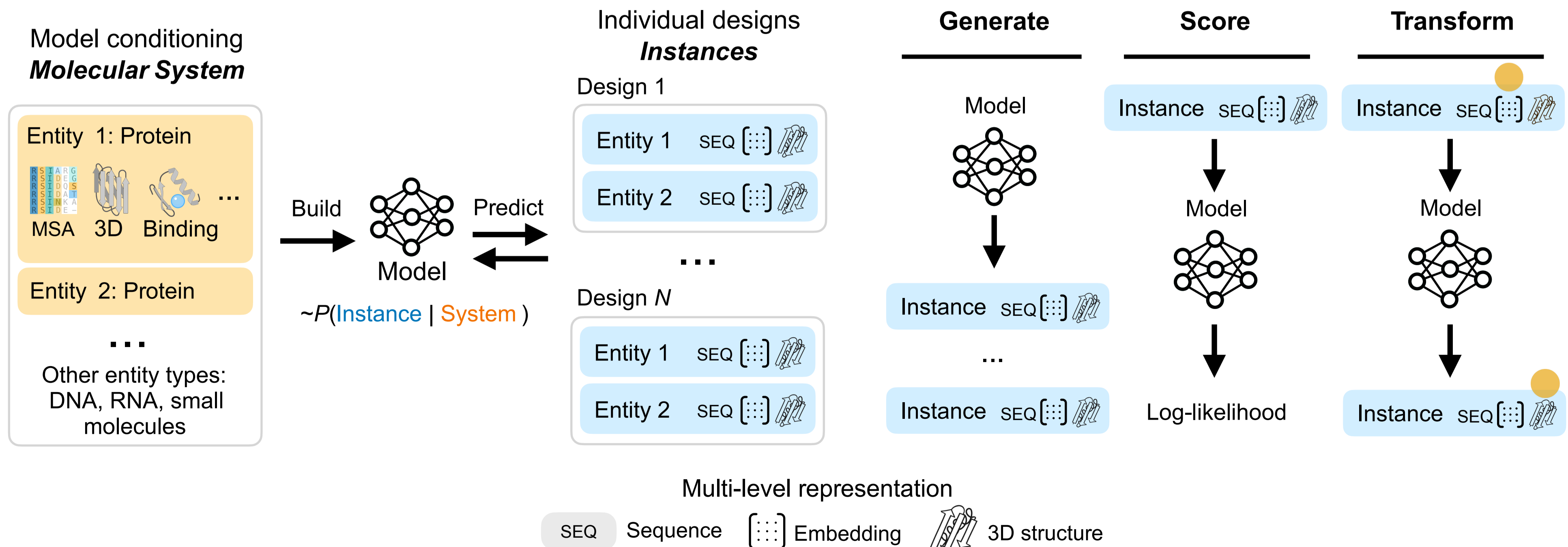
Our solution: a standardized framework

Biosequence design is framed as a **conditional modeling problem**



The user specifies a **system** with any known information, and models operate on system **instances**, which are realizations of the system.

All workflows are expressed through three composable operations: **generate**, **score** and **transform**.



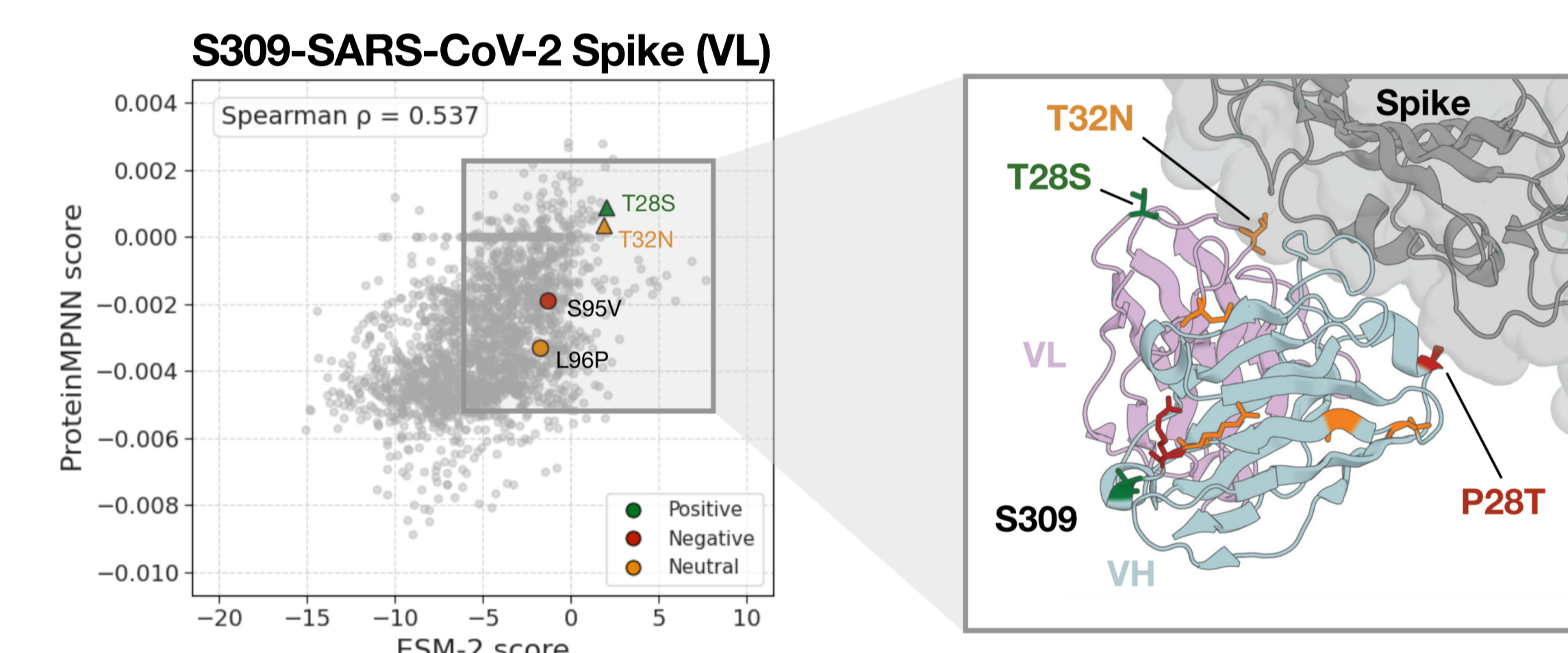
Case study I: antibody maturation

- We scored **single-residue substitutions** of clinically relevant antibodies with **ESM-2** and **ProteinMPNN**.
- We then compared scores with experimentally determined **antigen binding effects** for a panel of ESM-1-derived mutants (Hie *et al.*)

Mutation effects close to the antigen interface could **only be predicted with ProteinMPNN**, suggesting that models contain **complementary** information, easily accessible with evedesign.

Score from interface mutations

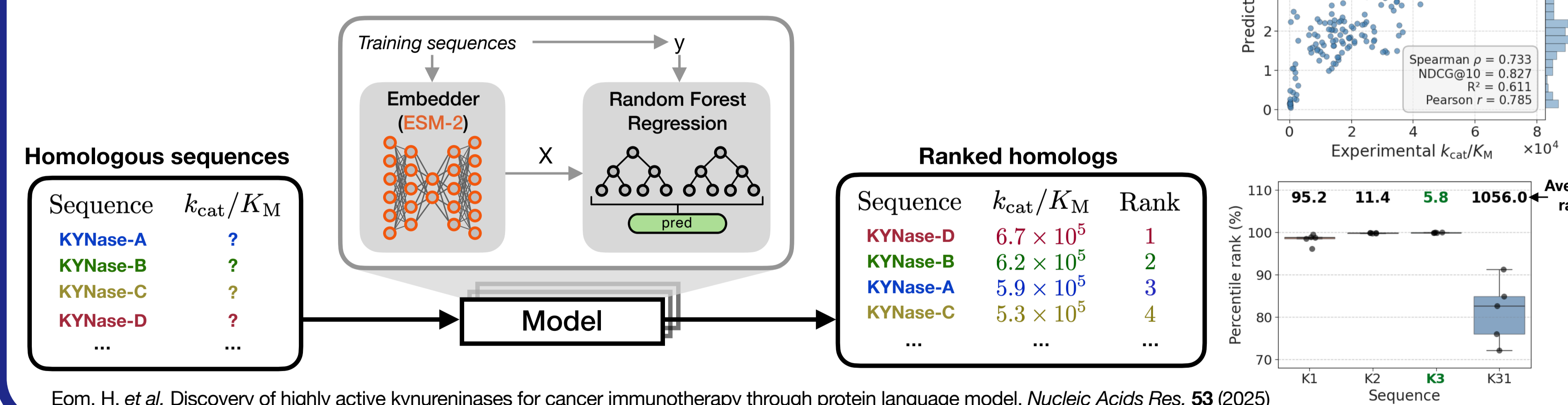
	ESM-2	MPNN	
100	99.0	98.6	T28S (S309, VL)
95	99.8	99.6	I96S (REGN, VL)
90	98.7	96.8	T32N (S309, VL)
85	99.9	80.0	M31S (mAb114, VH)
80	99.8	72.2	P28T (S309, VH)
75	98.9	65.8	S49Y (mAb114, VL)



Hie, B. L. *et al.* Efficient evolution of human antibodies from general protein language models. *Nat Biotechnol.* **42**, 275-283 (2024)

Case study II: enzyme discovery

- We trained an **ESM-2 embeddings-based regressor** to predict the k_{cat}/K_M of kynureninases (KYNases) matching an existing study (Eom *et al.*)
- We then **scored** a collection of **5,676 KYNase homologs**, consistently ranking the most efficient homolog (K3) *amongst the top 10* across random seeds.



Eom, H. *et al.* Discovery of highly active kynureninases for cancer immunotherapy through protein language model. *Nucleic Acids Res.* **53** (2025)

Towards a community-wide design and benchmarking suite

evedesign ProteinGym2

For developers: Integrate your model with evedesign and seamlessly **benchmark** it on a wide variety of protein design tasks across **>10M** mutation effects

For experimentalists: Build and test resulting pipelines with different models *in silico* before purchasing library

