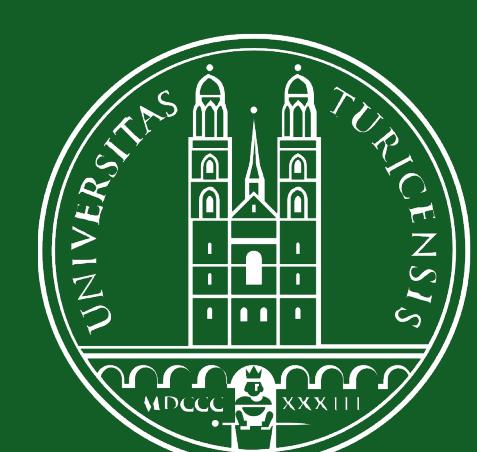


Generative model-assisted design of linear-peptide-recognizing armadillo repeat proteins

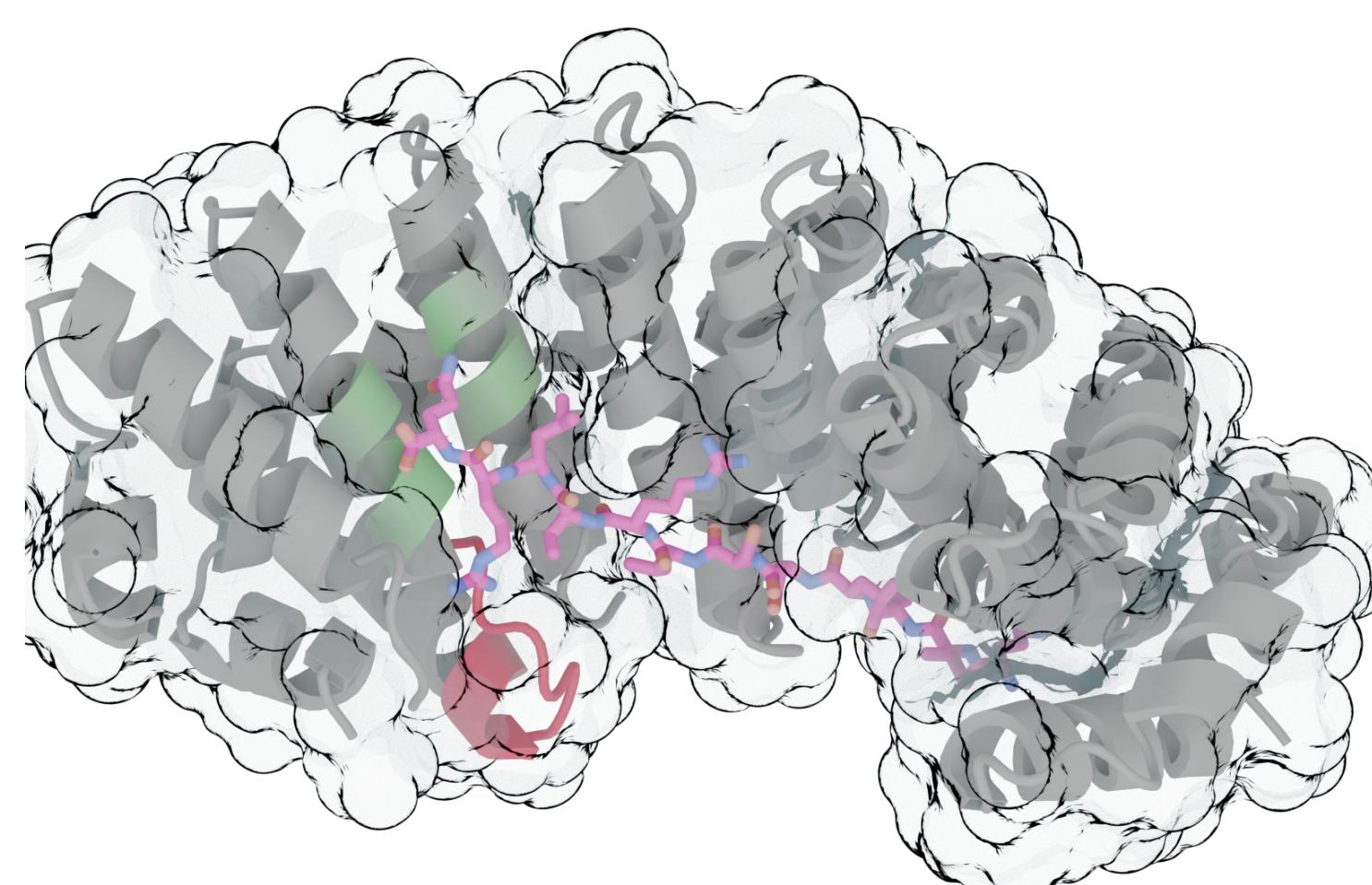


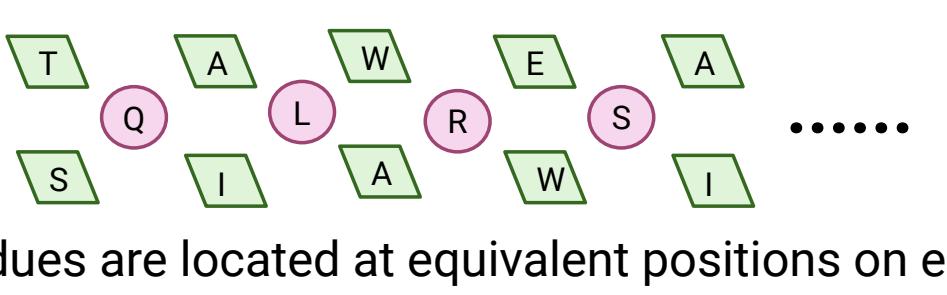
Universität
Zürich ^{UZH}

Songyuan Liu, Amedeo Caflisch, Andreas Plückthun
Department of Biochemistry, University of Zurich, Switzerland

life science zurich

ArmRP-Peptide Complex

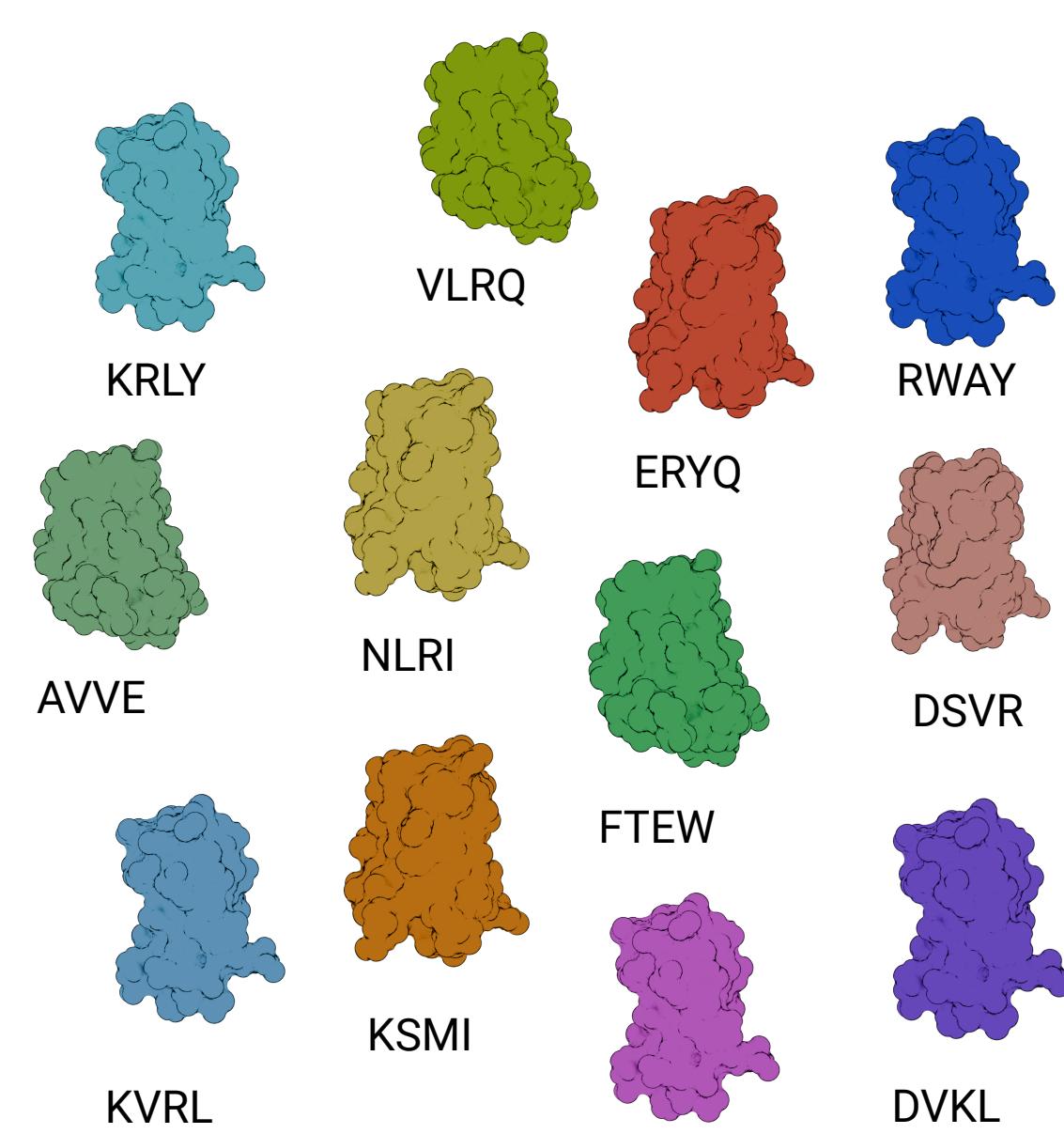


Backbone-fixed amino acid-recognition pocket:


Amino acid binding residues are located at equivalent positions on each repeat α -helix. The constrained relative distances between the target amino acid and the binding residues enable reuse of the same binding pocket across different module arrangements. Most of these pockets originate from the yeast surface display selection and will be adjusted through the computational design pipeline upon assembly.

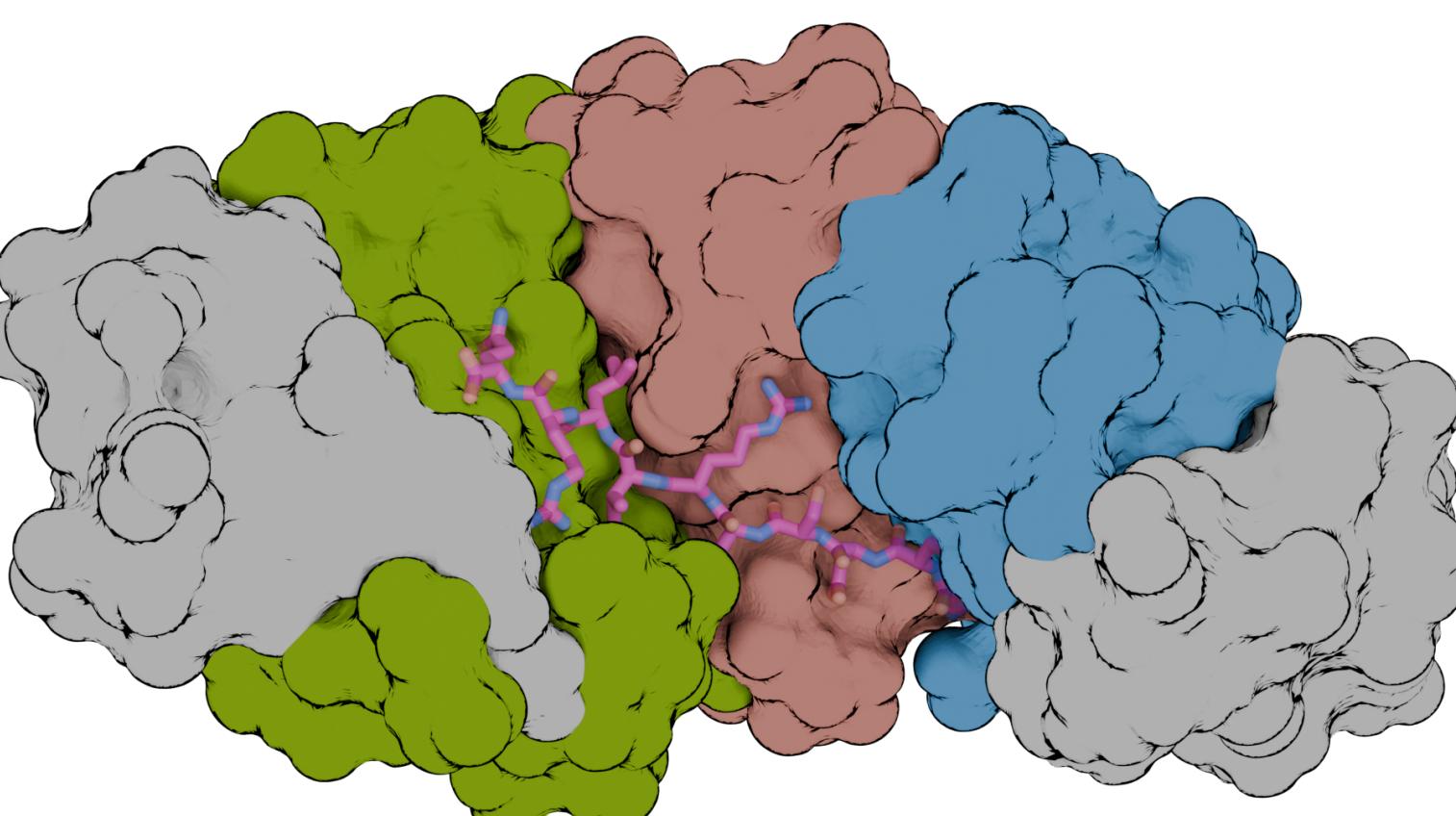
Flexible-backbone amino acid-recognition pocket:


The pockets with extended loops can be attached to any repeat modules and preserve their amino acid binding conformation. The specific loop conformations enhances the selectivity of pockets against off-target amino acids. A dedicated pipeline has been developed to design these de novo binding pockets.



Build repeat modules with different amino acid recognizing pockets enables the binding of arbitrary combination of amino acids.

Target Peptide
N.....KVRLDSVRVLRO.....C'
KVRL DSVR VLRO



Assembly of repeat modules supports the recognition of linear peptides with specific sequences

The final assembled dArmRPs specifically bind linear peptides with the intended sequences. The dArmRP-peptide complexes adopt an anti-parallel binding conformation, in which the N-cap of the dArmRP (left, gray) interacts with the C-terminus of the peptide (magenta), and conversely, the C-cap of the dArmRP engages with the N-terminus of the peptide.

INTRODUCTION

This research introduces a pioneering approach in protein engineering, focusing the design of Armadillo repeat proteins (ArmRPs) that specifically recognize linear peptides, enabling the detection of intrinsically disordered proteins. Leveraging the intrinsic modular repetitiveness of ArmRPs, the designed and engineered modules exhibit high transferability. These modules can be recombined in various arrangements to recognize new peptides while preserving the function of the specific amino acid binding pocket with minimal loss. This strategy eliminates the need for repetitive traditional screening, significantly reducing both time and experimental costs.

HIGHLIGHT

Advanced generative models, such as diffusion models and large language models informed by local structural and affinity data, enable the efficient creation of specific and reproducible binding pockets on known scaffolds or the optimization of binding interfaces between ArmRPs and peptides to enhance affinity and specificity. These computational designs often yield functional hits within a few experimental tests, substantially reducing the cost of building DNA libraries and labor associated with traditional display-based screening.

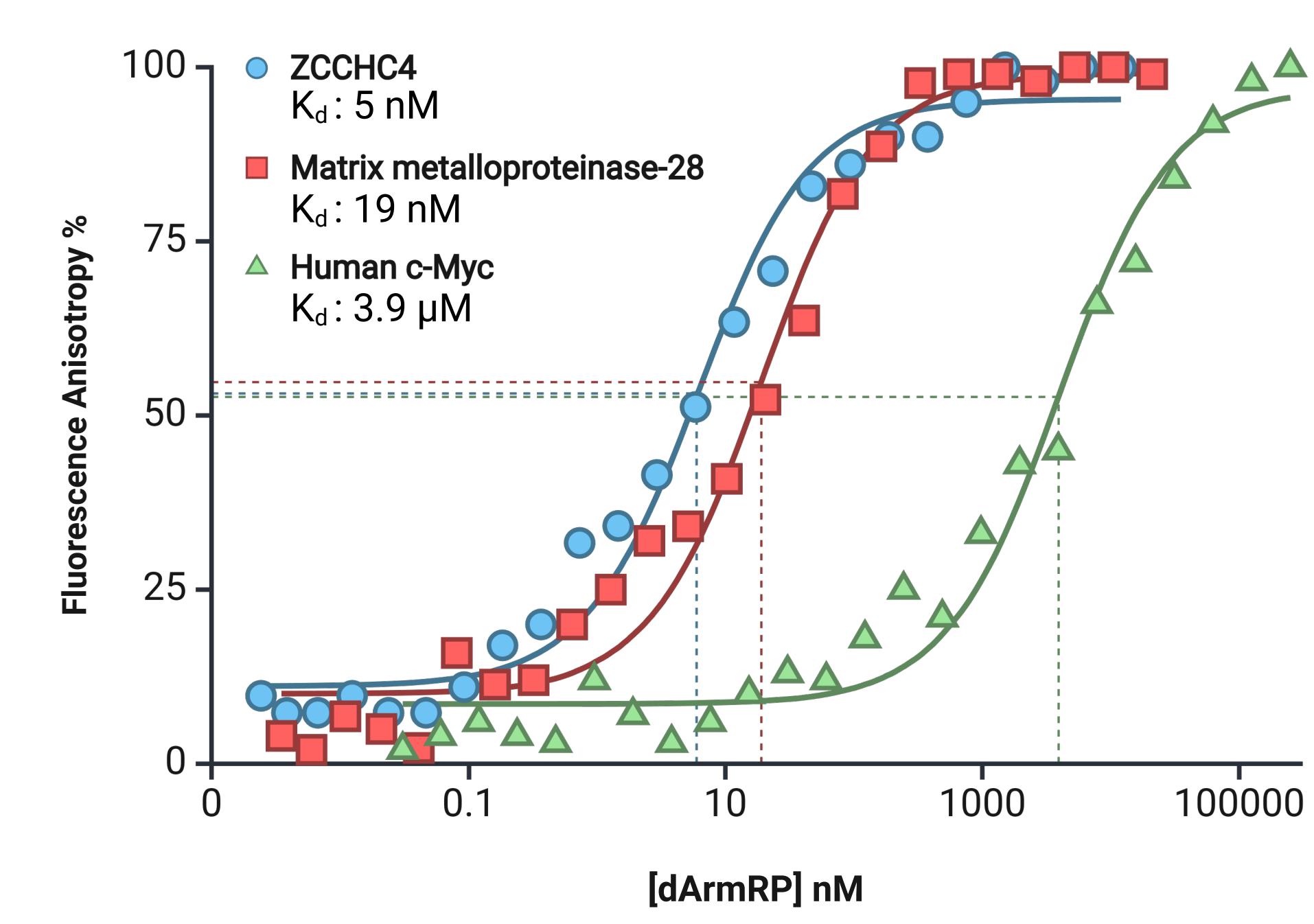


OBJECTIVE

We are developing an *in silico* pipeline comprising modules for binding motif generation, inverse co-folding sequence design, and binding complex structure prediction. Designed binders are selected based on prediction confidence metrics, geometric constraints, and coarse-grained force field energy. The selected candidates are expressed in *E. coli* and tested *in vitro* for their binding affinity to the target peptides.

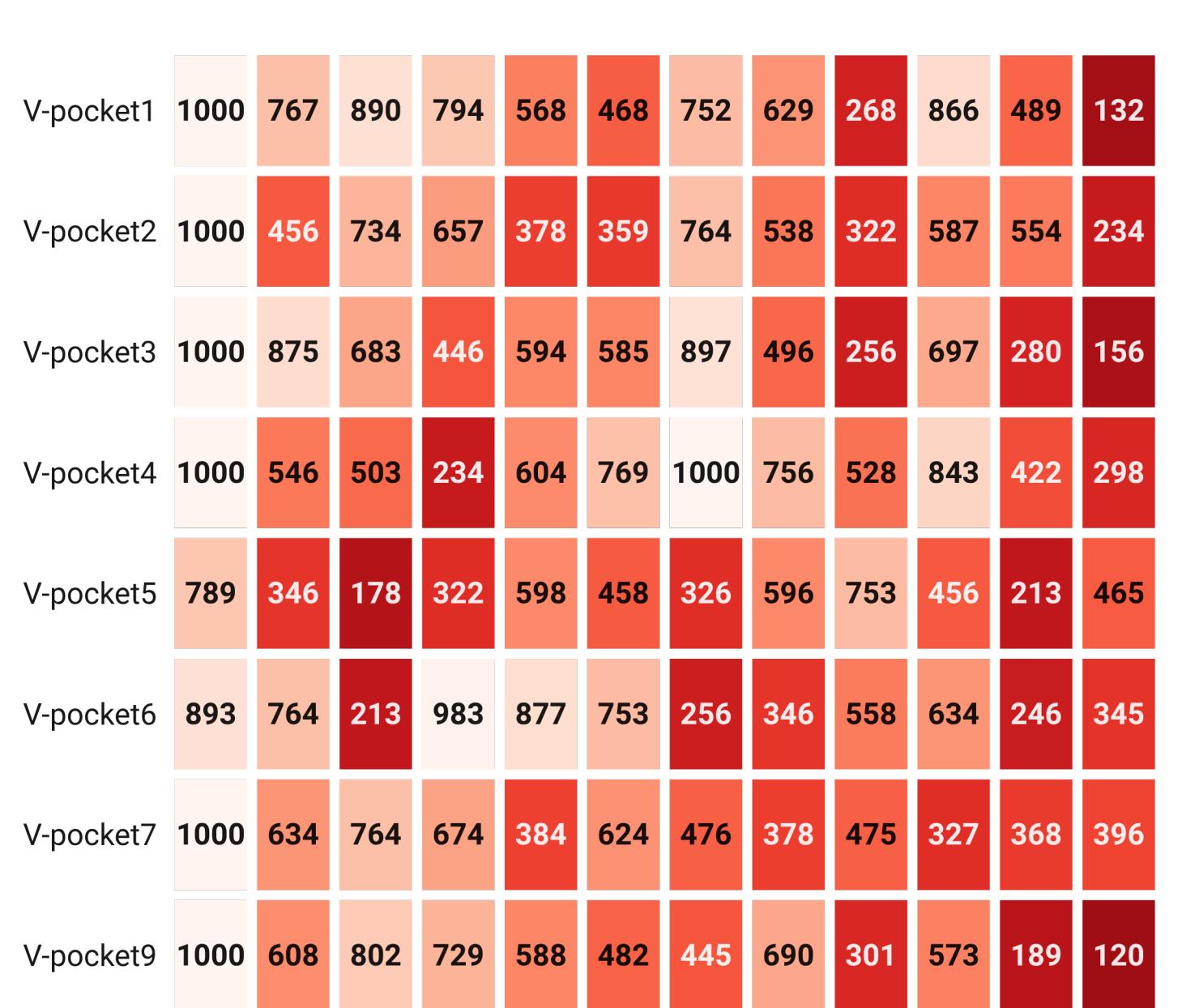
EXPERIMENTAL RESULT

Affinity between dArmRPs and Human Peptides



We optimized or selected various dArmRPs from the computational pipeline to bind peptides derived from human proteins. For certain targets, such as rRNA N(6)-adenosine-methyltransferase (ZCCHC4) and matrix metalloproteinase 28, the dArmRPs exhibited single- to double-digit nanomolar affinities. Other designs, such as those targeting a peptide from the human c-Myc protein, showed only micromolar affinities, indicating room for further improvement through iterative design.

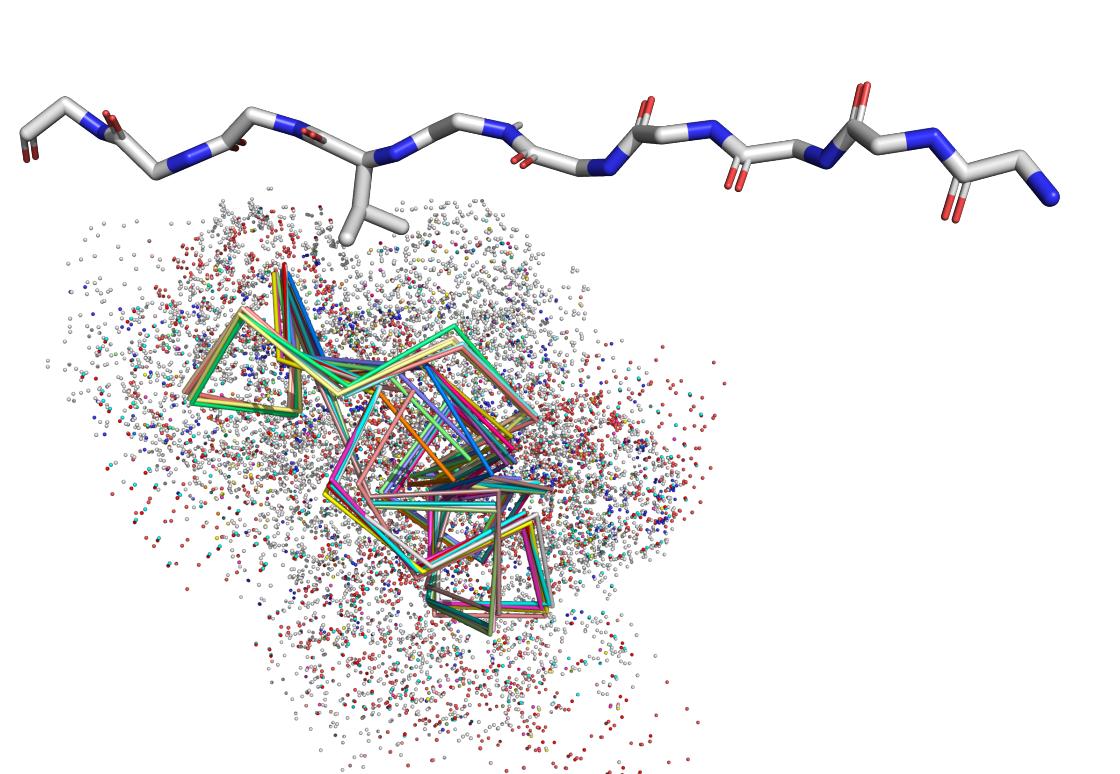
Affinity and Specificity Tests of single Pocket



This heat map summarizes the experimentally measured affinities and specificities of designed single-amino acid recognition pockets. The example shown here presents results for valine-binding pockets. Each row corresponds to an ArmRP variant containing a uniquely designed valine-binding loop at the same position, as illustrated in the Prediction step of the Computational Design section. The columns represent peptides with different amino acids substituted at the defined binding position (e.g., xxVxxxxxxxxx, xxRxxxxxxxxx), where the peptide containing valine is the intended target, and those with other residues serve as off-target controls. The labeled values indicate the experimentally measured K_d between each ArmRP and peptide. For a successful design, the ArmRP is expected to exhibit the lowest K_d for the valine-containing peptide. The maximum value on the heat map scale is capped at 1000 nM, though actual values may exceed this threshold.

COMPUTATIONAL DESIGN

Design

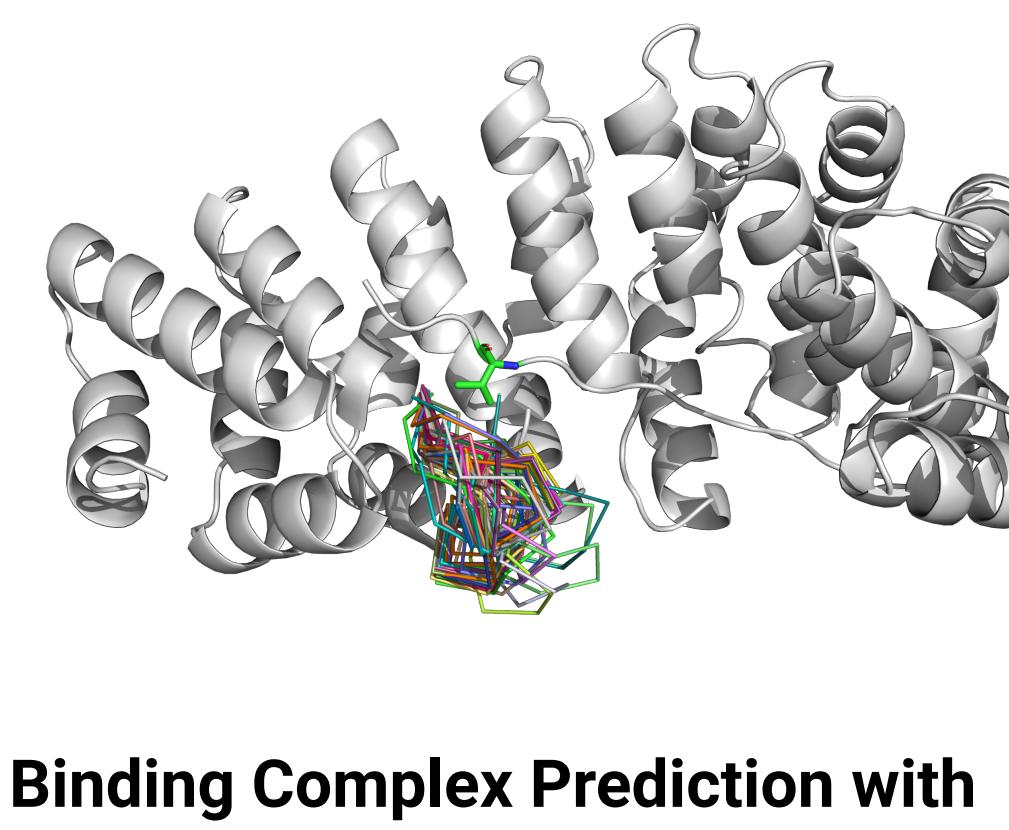


Pocket Backbone Generation with RFdiffusion

To expand the conformational sampling space for pockets with flexible backbones, RFdiffusion was employed as a binding motif generator to construct pockets ranging from six to twelve amino acids in length.

Sequence Design with ProteinMPNN/LigandMPNN/FAMPNN

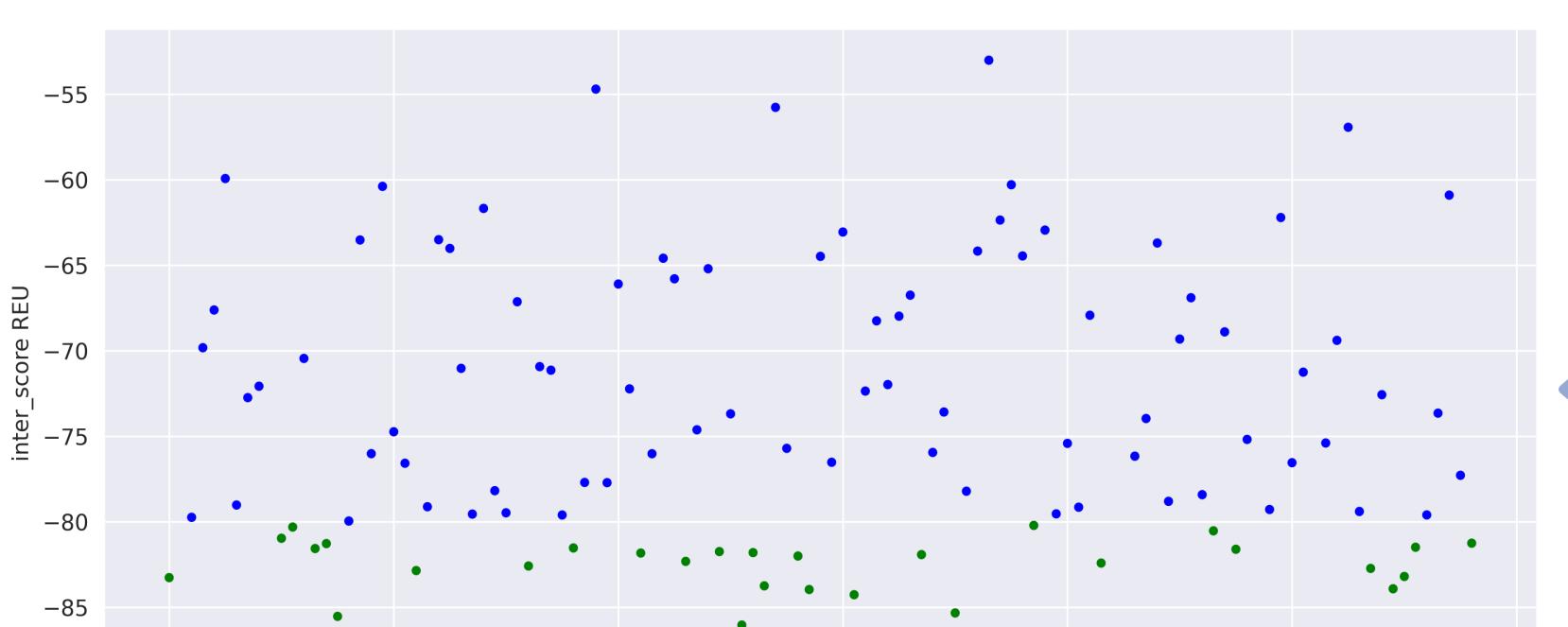
Sequences are designed either for pocket backbones generated by RFdiffusion or by redesigning residues on fixed-backbone pockets after assembly to refine binding interactions. To enhance sequence diversity, melting temperature and noise levels are manually adjusted.



Binding Complex Prediction with AlphaFold2/3

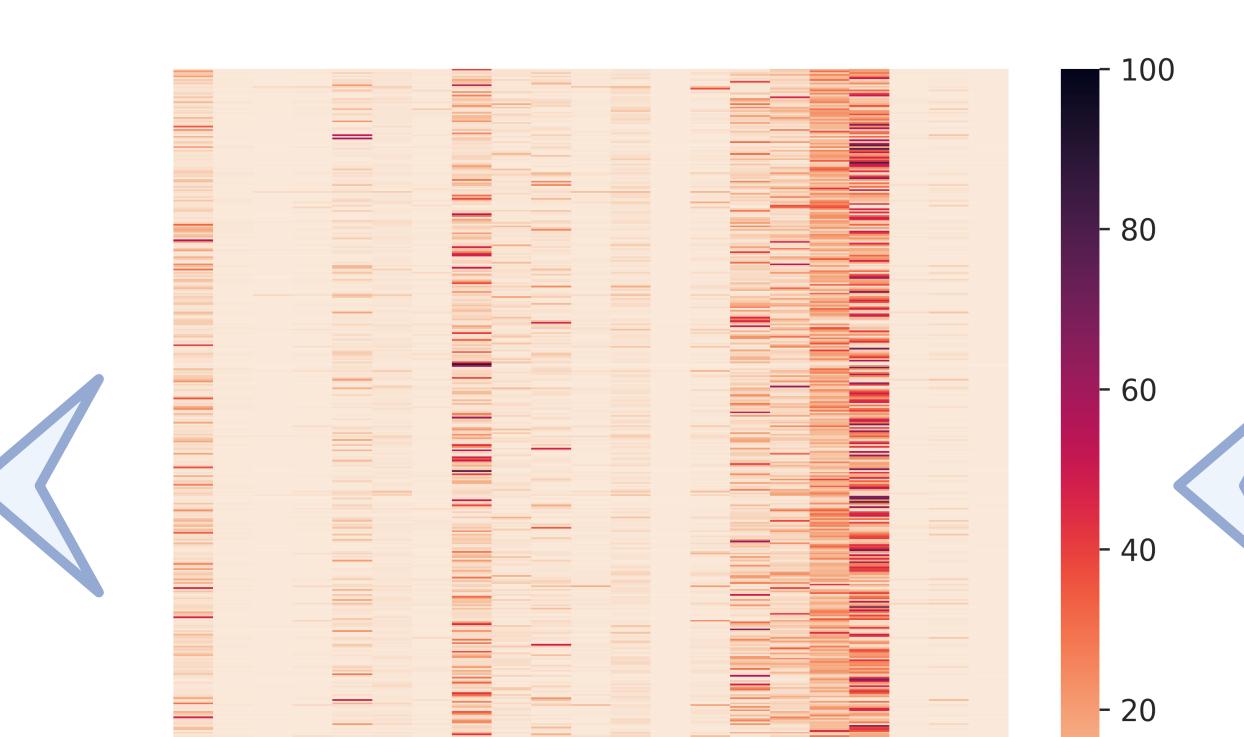
We employed the localcolafold with MSA calculated by mmseq2 as an alternative for AF2, and Proteinex developed by ByteDance as the AF3 replacement to increase the efficiency of the prediction. The models are predicted without templates.

Evaluation



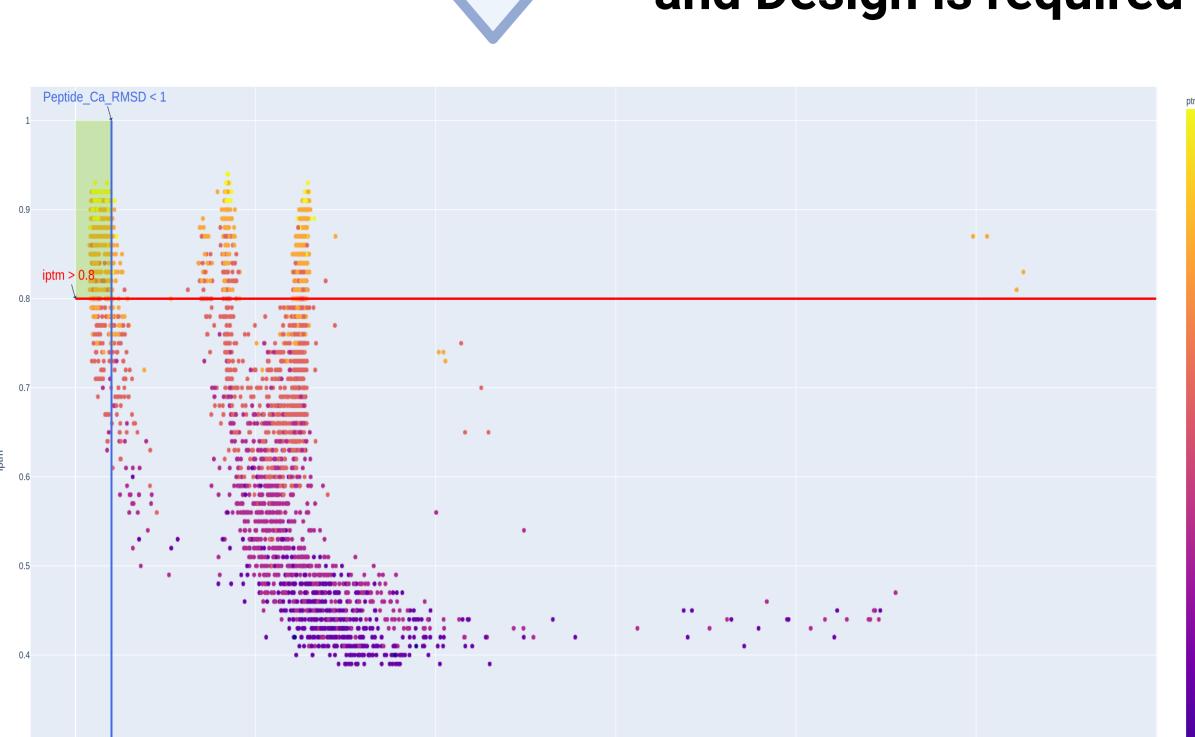
$$\Delta G_{interface} = \sum_i \Delta G_{i|Pockets} + \sum_j \Delta G_{j|Peptides}$$

We calculate the defined interface free energy with Rosetta energy function on defined residues in binding pockets and peptide. The models have been relaxed with Amber/Rosetta force field.



Specificity Control with Conditional Probability

The conditional probability calculated by ProteinMPNN has been applied as a control of selectivity for designed flexible pockets. Each row in the heat map represents the AA probability of target AA on the peptide bound with designed pockets.



Model Filter by ipTM and RMSD

The ipTM values have been extracted from predictions metrics. The RMSD here is the RMSD of peptide C_α after align dArmRP between predictions and crystal structures.

OUTLOOK

Current efforts focus on expanding the number of amino acid-recognition pockets and targeting more challenging epitopes. With a growing dataset of experimentally validated ArmRP-peptide binding pairs, we are fine-tuning large protein language models using our locally curated paired sequence database, aiming to generate ArmRP sequences conditioned on input peptide sequences.