GENERALIZED MULTI-STATE PROTEIN DESIGN WITH ALPHAFOLD3

Xiwei Cheng[1], **Pengkang Guo**[2], **Seonghwan Seo**[3], **Casper A. Goverde**[2], **Wengong Jin**[1]
[1] Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA
[2] École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[3] Department of Chemistry, KAIST, Daejeon, South Korea
{cheng.xiw,w.jin}@northeastern.edu,
{pengkang.guo,casper.goverde}@epfl.ch, {shwan0106}@kaist.ac.kr

ABSTRACT

While recent advances in deep learning have revolutionized single-state protein design, engineering proteins that can adopt multiple distinct structural states remains a fundamental challenge. In this work, we present a generalized framework for multi-state protein design leveraging the advanced all-atom co-folding capabilities of AlphaFold3. We first introduce AF3-CLUSTER, an evaluation framework that synergizes Multiple Sequence Alignment (MSA) clustering with AlphaFold3 to resolve complex structural ensembles across diverse biochemical contexts. Validation on a benchmark of ten multi-state proteins demonstrates that AF3-CLUSTER robustly recovers alternative conformations, particularly for binder-mediated transitions where previous single-chain methods often fail. Building on this, we propose AF3-MSD, an evolutionary framework that utilizes AlphaFold3 as a computationally efficient scoring proxy to guide the sculpting of multi-basin energy landscapes. AF3-MSD successfully drives candidates toward higher scores to design functional sequences for seven multi-state proteins, achieving structural agreement on par with natural multi-conformational proteins.

1 INTRODUCTION

In complex biological systems, conformational transitions between distinct protein states are fundamental to diverse biological processes, including transmembrane transport, enzyme catalysis, and signal transduction. Harnessing these transitions through multi-state protein design holds transformative potential for biological engineering, enabling applications such as programmable biosensors, targeted drug delivery, and synthetic molecular motors. However, designing for multiple states is significantly more challenging than traditional single-state design. It requires the precise sculpting of a global energy landscape to feature multiple low-energy basins with each corresponding to a functionally relevant state, while simultaneously suppressing undesired off-target minima. This creates an intrinsic tension among foldability, specificity, and controllability that remains unaddressed by standard single-structure design paradigms.

Recent research has explored various multi-state and switchable protein design strategies, achieving notable success in engineering hinge-like domain motions (Praetorius et al., 2023), fold-switching proteins (Guo et al., 2025), and conformationally bistable systems (Pillai et al., 2024). While these studies demonstrate remarkable control over conformational behavior, they often rely on system-specific heuristics and deep mechanistic insights. To provide a more generalizable framework, learning-based approaches such as ProteinMPNN Multi-State Design (MSD) (Dauparas et al., 2022), DynamicMPNN (Abrudan et al., 2025), and ProteinGenerator (Lisanza et al., 2025) have extended generative protein design to multi-state setting by optimizing sequences to be simultaneously compatible with multiple predefined target states. Nevertheless, several critical challenges remain. The scarcity of multi-conformational structural data, relative to the abundance of static protein structures, poses a challenge for models to robustly capture the sequence patterns that encode a multi-basin energy landscape and its associated conformational plasticity (Abrudan et al., 2025).

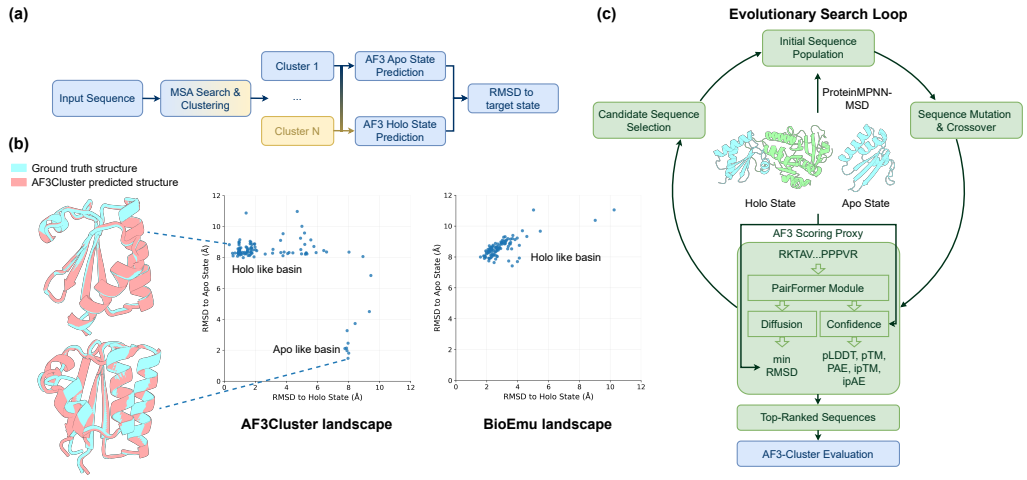

Figure 1: Overview of AF3-CLUSTER and AF3-MSD evolutionary design framework. **(a)** Evaluation pipeline of designed protein with AF3-CLUSTER. Each MSA cluster is independently used as input to AlphaFold3 to predict structures for the multiple target states. **(b)** Predicted landscapes of KaiB from AF3-CLUSTER and BioEmu. The RMSD between AF3-CLUSTER and BioEmu predicted structures to the target apo and holo states. **(c)** Overview of AF3-MSD framework. Initial sequence population generated by ProteinMPNN-MSD is iteratively updated through mutation, crossover, and selection. Candidate sequences evaluated using AF3 scoring proxy.

Furthermore, a comprehensive computational framework capable of generating representative protein ensembles across diverse biochemical contexts, such as ligand-bound states, is necessitated to validate designed proteins. Building on the principle that multiple sequence alignments (MSAs) encapsulate competing structural signals (Jumper et al., 2021), AF-Cluster (Wayment-Steele et al., 2024) demonstrated that MSA clustering can uncover alternative states in fold-switching proteins. More recently, generative approaches such as BioEmu (Lewis et al., 2025) have allowed for the scalable generation of equilibrium protein ensembles, capturing intricate domain motions and cryptic pocket formation. However, these methods primarily focus on protein monomers, often neglecting the crucial influence of binders that drive or stabilize conformational changes. To overcome these limitations, we propose AF3-CLUSTER, which integrates MSA clustering with the all-atom co-folding capabilities of AlphaFold3 (AF3) (Abramson et al., 2024). As illustrated in Figure 1b, AF3-CLUSTER accurately resolves the distinct apo- and holo-like basins of the metamorphic protein KaiB compared to recent ensemble generation model BioEmu. Validated on a benchmark of 10 representative multi-state proteins, AF3-CLUSTER successfully recovers alternative conformations in systems where transitions are mediated by binders, thereby establishing a robust evaluation framework for multi-state design.

While AF3-CLUSTER provides a reliable assessment, its high computational cost associated with ensemble sampling precludes its direct use in large-scale sequence optimization or de-novo design loops. To address this limitation, we introduce AF3-MSD, a evolutionary framework for multi-state protein design. Initialized with a diverse population of sequences from ProteinMPNN-MSD (Dauparas et al., 2022), AF3-MSD iteratively optimize candidate sequences by leveraging AF3 as a structure-based scoring proxy. At each iteration, AF3 provides **target-structure confidence score** and **state-specific structural agreement score** for each state, with full definitions provided in Section 3. Finally, top-scoring sequences are selected and validated using AF3-CLUSTER. We demonstrate the effectiveness of AF3-MSD across diverse multi-state targets involving protein binders (4 targets) or ligand binders (3 targets), showing that it successfully designs functional sequences with desired conformational transitionability.

2 AF3-CLUSTER: TOWARDS STRUCTURE ENSEMBLE GENERATION

AlphaFold3 (AF3) (Abramson et al., 2024) is the diffusion-based generative framework that enables the sampling of multiple plausible structures for a given input on diverse biomolecular inputs, such as protein monomers, protein-ligand complexes, and diverse multi-chain assemblies. Crucially, AF3 inherently distinguishes between apo (unbound) and holo (bound) states, capturing the confor-

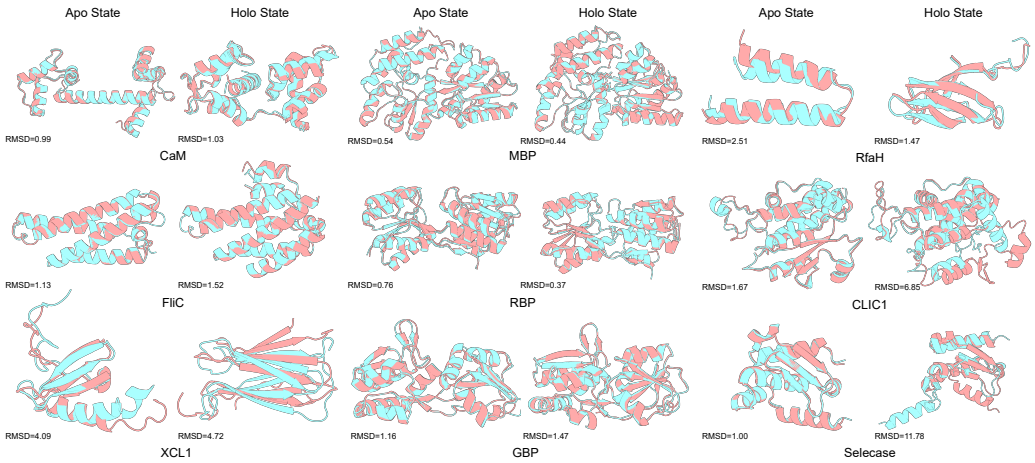

Figure 2: Visualization of ground truth state structures (blue) and AF3-CLUSTER predicted structures with minimum RMSD to the target states (red).

mational shifts induced by interactions with binders. By synergizing AF3's inherent generative capability with the MSA clustering strategy proposed by AF-Cluster (Wayment-Steele et al., 2024), AF3-CLUSTER maximizes the resolution and diversity of the generated structural ensembles.

Specifically, for each input protein sequence, we construct MSAs using ColabFold (Mirdita et al., 2022) and cluster them by edit distance using DBSCAN algorithm (Ester et al., 1996), thereby isolating sub-alignments that reflect distinct evolutionary trajectories. Each resulting MSA cluster is then independently provided as an input to AF3, enabling the model to explore specific conformational basins associated with those evolutionary signals. For systems involving binder-induced transitions as illustrated in Figure 2, AF3-CLUSTER explicitly incorporates binder information by leveraging the co-folding capabilities of AF3 to resolve state-specific ensembles.

To evaluate the designed sequences for multi-state protein design using AF3-CLUSTER, we quantify the structural similarity between the predicted ensemble members and given target state structures using RSMD. A sequence is considered to have successfully recovered the multi-state landscape if its predicted ensemble contains high-accuracy structures (low RMSD) for all targeted conformations. Detailed parameters for MSA construction and DBSCAN configurations used across our benchmarks are provided in Section A.

3 AF3-MSD: TOWARDS EVOLUTIONARY MULTI-STATE PROTEIN DESIGN

We formulate multi-state protein design as a sequence optimization problem over a global energy landscape defined by multiple target conformations. To navigate this space efficiently, we introduce AF3-MSD, a closed-loop evolutionary framework that utilizes AF3 as a proxy for the computationally intensive AF3-CLUSTER. The optimization process is initialized with a diverse population of candidate sequences generated by ProteinMPNN-MSD (Dauparas et al., 2022), providing a robust starting point based on established multi-state inverse folding priors.

At each iteration, the population is evolved through segment-based crossover and stochastic mutation operators. To balance exploration and exploitation, the subsequent generation is formed by selecting the top-performing sequences, retaining a random subset of the current population, and sampling new candidates via a softmax distribution over the fitness scores. Upon convergence, the highest-scoring sequence observed across all generations is selected as the final design.

To evaluate the fitness of each candidate, we define two complementary scoring terms:

- **Target-structure confidence score** assesses the consistency of a designed sequence with a target state by conditioning the AF3 confidence heads on the target structure. For apo states, we aggregate the predicted local distance difference test (pLDDT), predicted template modeling score (pTM), and predicted aligned error (pAE). For holo states, we additionally incorporate interface-specific metrics: the interface TM-score (ipTM) and interface aligned error (ipAE).

Table 1: Comparison of minimum RMSD (Å) between structures predicted by AFCluster and AF3-CLUSTERand target states for natural sequences.

State	Model	KaiB	CaM	FliC	XCL1	MBP	RBP	GBP
Apo state	AF-Cluster	1.42	4.40	1.89	3.92	1.27	0.80	3.69
	AF3-CLUSTER	1.47	0.99	1.13	4.09	0.54	0.76	1.16
Holo state	AF-Cluster	0.92	3.82	7.83	6.95	2.35	0.52	1.68
	AF3-CLUSTER	0.32	1.03	1.52	4.72	0.44	0.37	1.47

Table 2: Comparison of minimum RMSD (Å) between AF3-CLUSTER predicted structures and target states for natural sequences and AF3-MSD designed candidates. Results for AF3-MSD represent the average of the top 10 scoring sequences.

State	Method	KaiB	CaM	FliC	XCL1	MBP	RBP	GBP
Apo state	Natural Sequences	1.47	0.99	1.13	4.09	0.54	0.76	1.16
	AF3-MSD	1.31	1.01	2.01	4.19	0.91	0.77	3.22
Holo state	Natural Sequences	0.32	1.03	1.52	4.72	0.44	0.37	1.47
	AF3-MSD	0.98	0.89	1.14	4.85	0.70	0.44	2.07

- **State-specific structural agreement score** quantifies the sequence's propensity to fold into the desired conformations. It is defined as the minimum RMSD between a set of stochastic AF3 structure predictions and the corresponding target-state coordinates.

Additional details for evolutionary search and scoring are provided in Section B.

4 EXPERIMENTS

4.1 BENCHMARK DESIGN

We constructed a benchmark composed of ten multi-state proteins, including three intrinsic metamorphic proteins and seven proteins whose alternative conformations (holo states) are stabilized or induced by specific binders. Following Wayment-Steele et al. (2024), the metamorphic category includes **RfaH**, **Selecase**, and **CLIC1**. To capture binder-mediated transitions, we incorporate four protein-binding systems (**XCL1**, **KaiB**, **FliC**, **CaM**), and three ligand-binding systems (**MBP**, **RBP**, **GBP**). Details of each protein system are provided in Section C.

4.2 EVALUATION OF AF3-CLUSTER FOR MULTI-STATE STRUCTURE PREDICTION

To assess the ability of AF3-CLUSTER to recover multi-state conformational landscapes, we first evaluated its performance on the natural sequences within our benchmark. For each protein, we calculated RMSD between the predicted ensembles and the ground-truth structures for both target states. As illustrated in Figure 2, AF3-CLUSTER successfully captures both the apo and holo conformations for systems where the transition is stabilized by a binder. In the case of **XCL1**, while the model predicts distinct conformational states, the loop regions exhibit higher RMSD values, likely reflecting the intrinsic flexibility of these parts. For the set of intrinsic metamorphic proteins, our results are similar to those of the original AF-Cluster. While the model successfully predicts the distinct states of **RfaH**, it fails to recover the holo states of **Selecase** and **CLIC1**, consistently biased toward the apo-like conformations. The generated conformational landscapes of these intrinsic metamorphic proteins are provided in Section D.1 for clear interpretation.

In Table 1, we further compared AF3-CLUSTER with the AF-Cluster using AlphaFold2 (Jumper et al., 2021) specifically on binder-mediated proteins. Generally, AF3-CLUSTER achieves comparable or lower RMSD values, with particularly notable improvements in the accuracy of holo state predictions. These observations suggest that AF3-CLUSTER is effective for predicting multi-state structures when the alternative state is stabilized by a binder, whereas it may struggle with certain intrinsic metamorphic transitions. Consequently, we focus our subsequent design experiments on multi-state proteins involving binder-mediated stabilization.

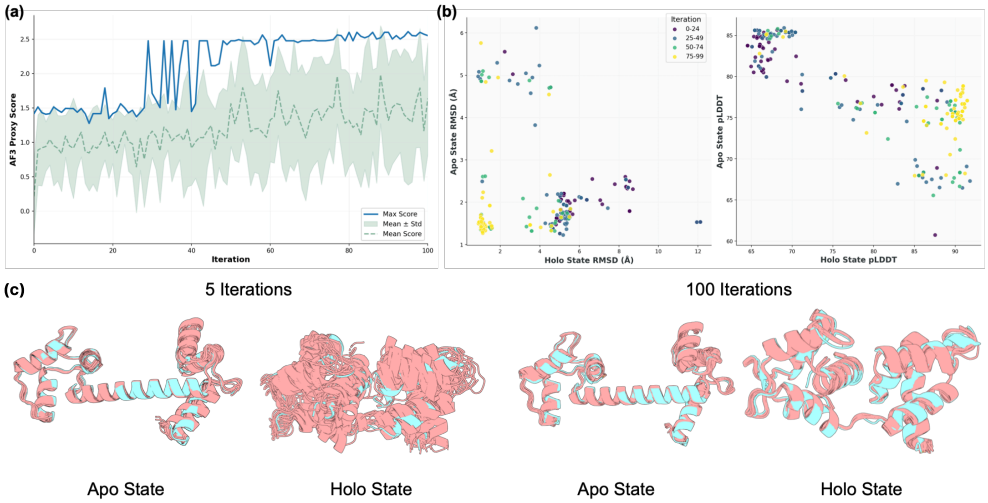

Figure 3: Optimization trend of CaM with AF3-MSD. **(a)** Total score across iterations. **(b)** RMSD and pLDDT Pareto boundary formed by top scoring sequences at 25 iteration intervals. **(c)** Structural ensembles of top-scoring CaM sequences predicted by AF3-CLUSTER at 5 and 100 iterations.

4.3 MULTI-STATE PROTEIN DESIGN WITH AF3-MSD

Setup. For each target protein, we initialized the evolutionary search with 100 candidate sequences generated by ProteinMPNN-MSD and evolved the population for 100 iterations. The resulting designs were then ranked by their final AF3 proxy scores, and the top 10 sequences for each target were selected for post-hoc validation using the AF3-CLUSTER ensemble generation pipeline. Detailed hyperparameters for the evolutionary operators and AF3 scoring are provided in Section B.3.

Results. To characterize the optimization dynamics of AF3-MSD, we select CaM as a representative case study. As shown in Figure 3a, the AF3 proxy scores of the population exhibit a rapid increase during the initial 50 iterations before reaching a plateau as the search converges. This improvement is further reflected in the Pareto frontier between apo and holo states' predicted RMSD and pLDDT (Figure 3b), which consistently shifts toward higher confidence and lower structural deviation across both states. We further observed that sequences from the early stages of optimization often fail to capture the desired multi-state behavior. As illustrated in Figure 3c, these early candidates, which largely inherit the characteristics of the initial ProteinMPNN-MSD population, tend to collapse into the apo-state basin when evaluated with AF3-CLUSTER. In contrast, sequences optimized for 100 iterations robustly recover both the apo and holo conformations, demonstrating the necessity of the evolutionary refinement process in sculpting a dual-basin energy landscape.

The overall performance of AF3-MSD across our benchmark is summarized in Table 2. On average, the designed sequences achieve structural agreement with the target states that is on par with natural sequences. Detailed ensemble visualizations for all benchmark targets are provided in Section D.2. Further analysis of the influence of initialization and iteration length is presented in Section D.3.

5 CONCLUSIONS

In this work, we presented a generalized framework for multi-state protein design by leveraging the advanced generative capabilities of AlphaFold3. Our proposed framework, AF3-MSD, employs an evolutionary optimization strategy that utilizes AF3 as a computationally efficient scoring proxy to guide the sculpting of multi-basin energy landscapes. To evaluate design fidelity, we introduced AF3-CLUSTER, a protein ensemble generation method that synergizes MSA clustering with all-atom co-folding to resolve complex conformational landscapes across diverse biochemical contexts.

Our experimental results demonstrate that the combination of AF3-MSD and AF3-CLUSTER successfully designs functional sequences for seven binder-mediated multi-state systems, achieving structural agreement with target states on par with natural metamorphic proteins. We further show that the evolutionary refinement process is essential for overcoming the limitations of initial inverse-

folding populations, which often fail to capture the necessary conformational plasticity. By shifting from system-specific heuristics toward a high-fidelity, structure-aware design paradigm, this framework establishes a robust foundation for engineering the next generation of multi-state proteins. Future research will focus on integrating experimental feedback to further refine the predictive accuracy and generalizability of the design loop.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Alex Abrudan, Sebastian Pujalte Ojeda, Chaitanya K Joshi, Matthew Greenig, Felipe Engelberger, Alena Khmelinskaia, Jens Meiler, Michele Vendruscolo, and Tuomas PJ Knowles. Multi-state protein design with dynamicmpnn. *arXiv preprint arXiv:2507.21938*, 2025.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

Patrick Bryant and Frank Noé. Structure prediction of alternative protein conformations. *Nature Communications*, 15(1):7328, 2024.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

Justas Dauparas, Gyu Rie Lee, Robert Pecoraro, Linna An, Ivan Anishchenko, Cameron Glasscock, and David Baker. Atomic context-conditioned protein sequence design using ligandmpnn. *Nature Methods*, pp. 1–7, 2025.

Diego Del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with alphafold2. *Elife*, 11:e75751, 2022.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pp. 226–231, 1996.

Amy B Guo, Deniz Akpinaroglu, Christina A Stephens, Michael Grabe, Colin A Smith, Mark JS Kelly, and Tanja Kortemme. Deep learning–guided design of dynamic proteins. *Science*, 388 (6749):eadr7094, 2025.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, et al. Simulating 500 million years of evolution with a language model. *Science*, 387(6736):850–858, 2025.

Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Alphafold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845*, 2024.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Yogesh Kalakoti and Björn Wallner. Afsample2 predicts multiple conformations and ensembles with alphafold2. *Communications Biology*, 8(1):373, 2025.

Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew YK Foong, Victor García Satorras, Osama Abdin, Bastiaan S Veeling, Iryna Zaporozhets, et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761): eadv9817, 2025.

Sidney Lyayuga Lisanza, Jacob Merle Gershon, Samuel WK Tipps, Jeremiah Nelson Sims, Lucas Arnoldt, Samuel J Hendel, Miriam K Simma, Ge Liu, Muna Yase, Hongwei Wu, et al. Multistate and functional protein design using rosettafold sequence space diffusion. *Nature biotechnology*, 43(8):1288–1298, 2025.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature methods*, 2022.

Arvind Pillai, Abbas Idris, Annika Philomin, Connor Weidle, Rebecca Skotheim, Philip JY Leung, Adam Broerman, Cullen Demakis, Andrew J Borst, Florian Praetorius, et al. De novo design of allosterically switchable protein assemblies. *Nature*, 632(8026):911–920, 2024.

Florian Praetorius, Philip JY Leung, Maxx H Tessmer, Adam Broerman, Cullen Demakis, Acacia F Dishman, Arvind Pillai, Abbas Idris, David Juergens, Justas Dauparas, et al. Design of stimulus-responsive two-state hinge proteins. *Science*, 381(6659):754–760, 2023.

Richard W Shuai, Tianyu Lu, Subhang Bhatti, Petr Kouba, and Po-Ssu Huang. Ensemble-conditioned protein sequence design with caliby. *bioRxiv*, pp. 2025–09, 2025.

Hannah K Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and alphafold2. *Nature*, 625(7996):832–839, 2024.

A AF3-CLUSTER SETTINGS

Following the MSA clustering strategy described in AF-Cluster (Wayment-Steele et al., 2024), we utilize the DBSCAN algorithm (Ester et al., 1996) to partition multiple sequence alignments into sub-alignments reflecting distinct evolutionary signals. The ϵ parameter for DBSCAN is optimized for each sequence by performing a grid search over the range $[3, 20]$, selecting the value that maximizes the total number of detected clusters based on a representative subset (25%) of the MSA. For each resulting cluster, AlphaFold3 (Abramson et al., 2024) is run with 10 recycling steps and 5 stochastic samples to resolve the associated conformational basin.

B AF3-MSD PIPELINE

Algorithm 1 AF3-MSD: Evolutionary Multi-State Protein Design

Input: Target states $\{S_1, \ldots, S_k\}$, iterations T, population size N
Output: Best designed sequences $\mathbf{x}^*$
$\mathcal{P}_0 \leftarrow$ Initialize population of size N using ProteinMPNN-MSD
for $t = 0$ **to** $T - 1$ **do**
  Evaluate $\mathcal{P}_t$:
  for each sequence $\mathbf{x} \in \mathcal{P}_t$ **do**
    for each state S_i **do**
      Generate 5 stochastic AF3 predictions $\hat{Y}_i = \{\hat{y}_{i,1}, \ldots, \hat{y}_{i,5}\}$
      Compute $RMSD_i \leftarrow \min_j \text{RMSD}(\hat{y}_{i,j}, S_i)$
      Compute confidence metrics $\{pLDDT, pAE, ipTM, ipAE\}_i$
      $f_i(\mathbf{x}) \leftarrow \left(\frac{pLDDT}{100} + pTM - \frac{pAE}{31} - \frac{RMSD}{10}\right)_i$
      if binder$_i$ **then**
        $f_i(\mathbf{x}) \leftarrow f_i(\mathbf{x}) + \left(0.2 \cdot ipTM - 0.3 \cdot \frac{ipAE}{31}\right)_i$
  Normalize $\{f_i(\mathbf{x})\}$ across $\mathcal{P}_t$ using per-state z-score $\rightarrow \hat{f}_i(\mathbf{x})$
  $Score(\mathbf{x}) \leftarrow \sum_{i=1}^{k} \hat{f}_i(\mathbf{x})$
  Evolve next generation candidates $\mathcal{O}_t$:
  while $|\mathcal{O}_t| <$ offspring_size **do**
    Parent $\mathbf{x}_p \leftarrow$ Sample from $\mathcal{P}_t$ via Softmax$(Score)$
    if Rand$() < 0.8$ **then** ▷ Crossover
      $\mathbf{x}_q \leftarrow$ Sample another parent from $\mathcal{P}_t$
      $\mathbf{x}_{new} \leftarrow$ SegmentCrossover$(\mathbf{x}_p, \mathbf{x}_q)$
    else ▷ Mutation
      $\mathbf{x}_{new} \leftarrow$ SegmentMutation$(\mathbf{x}_p)$
    Add $\mathbf{x}_{new}$ to $\mathcal{O}_t$
  Evaluate $\mathcal{O}_t$ using the same scoring procedure as above
  Form $\mathcal{P}_{t+1}$:
  $\mathcal{P}_{t+1} \leftarrow$ Top 15 sequences from $\mathcal{O}_t$
  $\mathcal{P}_{t+1} \leftarrow \mathcal{P}_{t+1} \cup \{5$ random sequences from $\mathcal{O}_t\}$
  $\mathcal{P}_{t+1} \leftarrow \mathcal{P}_{t+1} \cup$ sample remaining $(N - 20)$ via Softmax$(Score)$ from $\mathcal{O}_t$
return $\mathbf{x}^* = \arg\max Score$ observed across all generations

B.1 AF3 PROXY SCORE COMPUTATION

To ensure that various confidence and structural metrics are on comparable scales, we normalize pLDDT by 100, RMSD by 10, and pAE and ipAE by 31. The composite score for each state is then calculated as a weighted sum: we assign coefficients of +1.0 to pLDDT and pTM, and -1.0 to normalized pAE and RMSD, while using +0.2 for ipTM and -0.3 for ipAE to consider interface-specific accuracy in holo states. Since score distributions can vary significantly between different conformational states, we apply per-state z-score normalization across the population before aggregation: $z_i = (x_i - \mu_i)/\sigma_i$, where x_i is the raw score for state i, and μ_i and σ_i are the population-wide mean and standard deviation for that state, respectively. The final AF3 proxy score is the sum of these normalized components across all target states.

B.2 Evolutionary Operators

The evolutionary optimization proceeds by selecting parent sequences with a probability proportional to the softmax of their AF3 proxy scores. New candidates are generated through crossover and mutation operators. With a probability of 0.8, we perform segment-based crossover by randomly selecting 1 to 3 segments (each with a length between 3 and 12 residues) from a second parent and grafting them onto the first. Alternatively, with a probability of 0.2, we apply stochastic mutations to 1 to 2 randomly selected segments of length 3 to 8. Following these operations, the resulting offspring are evaluated, and the next generation's population is formed by retaining the top 15 scoring sequences, incorporating 5 randomly sampled sequences to maintain diversity, and filling the remaining slots through softmax-based sampling from the candidate pool. The complete optimization pipeline for AF3-MSD is summarized in Algorithm 1.

B.3 Inference Settings and Hyperparameters

Throughout the evolutionary optimization, AF3 structure predictions are generated using 10 recycling steps. For each target state, we perform 5 stochastic samples with different random seeds and compute the minimum RMSD relative to the corresponding reference coordinates. To ensure the model receives a consistent signal regarding the desired multi-state transitions, the binder structures from the reference holo states are provided as explicit templates to AF3. This setup enables the confidence and error metrics to primarily reflect the conformational compatibility of the designable protein chain within each biochemical context.

C Benchmark

The target proteins and corresponding PDB IDs for distinct target states in our benchmark are summarized in Table 3.

For **CaM**, we specifically model the transition between the Ca^{2+}-free state and the peptide-bound state stabilized by the IQ motif-containing protein IQCG.

Table 3: Summary of target proteins and corresponding PDB IDs for target states in the benchmark. For protein-bound and ligand-bound proteins, state 1 and 2 indicate apo and holo states, respectively.

Category	**Protein**	**Full Name**	**State1**	**State2**
Metamorphic	RfaH	Transcription antitermination protein	5OND	2LCL
	CLIC1	Chloride intracellular channel protein	1K0N	1RK4
	Selecase	Archaeal metallopeptidase from *M. jannaschii*	4QHF	4QHH
Protein-bound	XCL1	Human cytokine	1J9O	2JP1
	FliC	Flagellin	1ORJ	1ORY
	CaM	Calmodulin	1CLL	4LZX
	KaiB	Cyanobacterial circadian clock protein	2QKE	5JWO
Ligand-bound	MBP	Maltose-binding protein	1OMP	1ANF
	RBP	Ribse-binding protein	1URP	1DRK
	GBP	Galactose-binding protein	1GGG	1WDN

D ADDITIONAL RESULTS

D.1 AF3-CLUSTER ANALYSIS

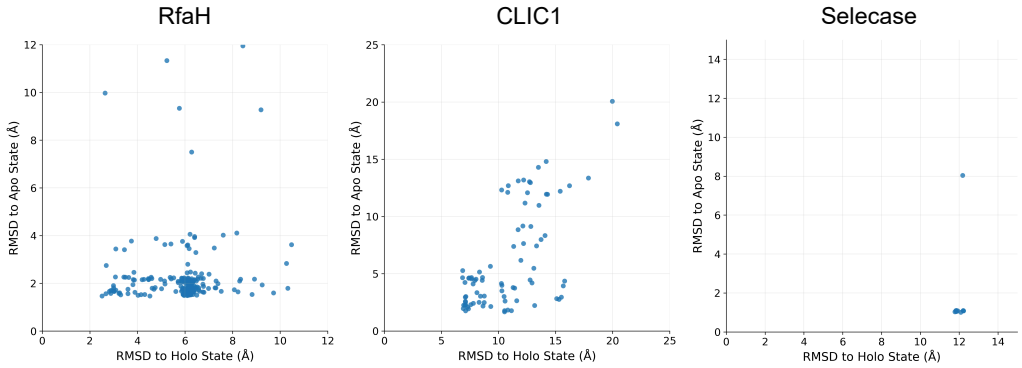

Figure 4: AF3-CLUSTER predicted landscapes of intrinsic metamorphic proteins RfaH, CLIC1, and Selecase.

We further analyzed the predicted landscapes of intrinsic metamorphic proteins: RfaH, CLIC1, and Selecase. The performance of AF3-CLUSTER is generally consistent with that of the original AF-Cluster on monomeric proteins. Specifically, both AF3-CLUSTER and AF-Cluster successfully recover both states for RfaH but fail to resolve the holo states for CLIC1 and Selecase. As shown in Figure 4, the predicted conformational landscapes for CLIC1 and Selecase are heavily biased toward their apo states, with minimum RMSD values to the holo states remaining around 7 Å and 12 Å, respectively. Because the performance of AF3-CLUSTER on intrinsic metamorphic transitions is fundamentally limited by the underlying structural foundation model's capability to capture alternative states from MSA signals alone, we focused the design scope of AF3-MSD on binder-mediated multi-state systems.

D.2 VISUALIZATION OF AF3-MSD DESIGNED PROTEINS

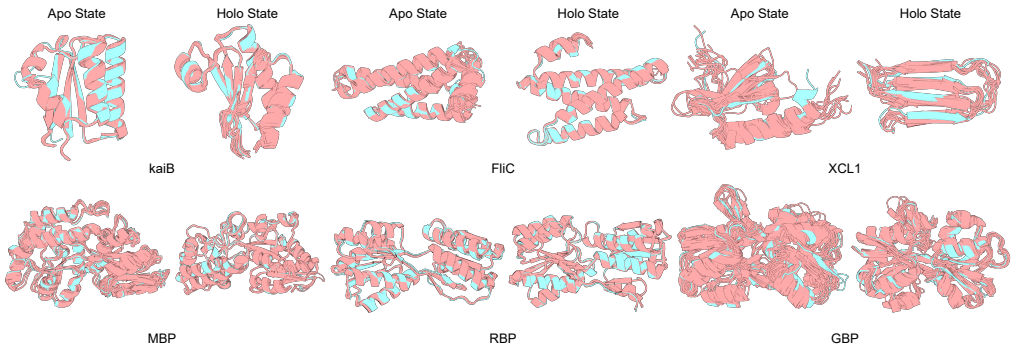

Figure 5: Visualization of AF3-CLUSTER predicted structure for the top-scoring sequences across benchmark proteins.

To qualitatively assess the design performance of AF3-MSD, we visualized the AF3-CLUSTER-predicted structural ensembles for the top-scoring sequences across benchmark proteins. For visual clarity, binding partners are omitted in these representations. As shown in Figure 5, the designed sequences successfully recover both targeted conformational states, demonstrating the robust performance achieved by AF3-MSD.

D.3 Ablation Studies

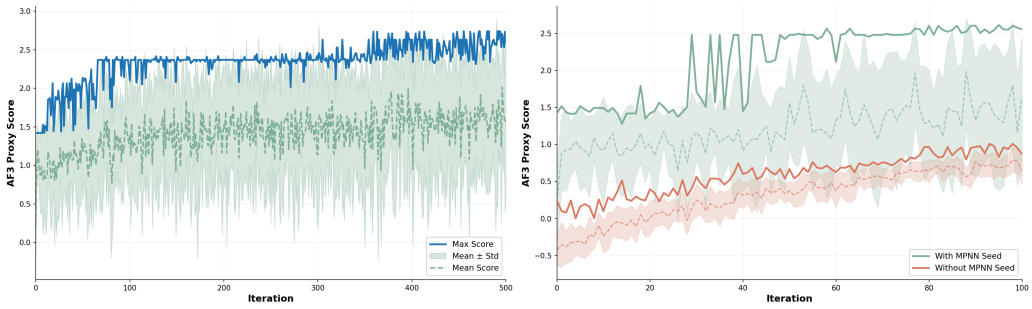

Figure 6: Ablation analysis of AF3-MSD optimization on CaM. The line plots show the impact of (left) the number of evolutionary search iterations and (right) population initialization using ProteinMPNN-MSD candidates, measured by AF3 proxy scores.

To evaluate the impact of ProteinMPNN-MSD initialization and evolutionary search iterations on the performance of AF3-MSD, we conducted ablation studies using CaM as a test case. First, we extended the evolutionary search to 500 iterations to examine the convergence behavior. As shown in Figure 6 (left), AF3 proxy scores increase rapidly during the initial 100 iterations, followed by a period of steady refinement. Notably, the maximum score continues to rise even after 400 iterations, suggesting that further optimization might yield additional improvements in design quality. Second, we compared the optimization performance of AF3-MSD with and without initialization from ProteinMPNN-MSD. As illustrated in Figure 6 (right), runs initialized with ProteinMPNN-MSD consistently achieve higher AF3 proxy scores compared to those starting from random sequences. It takes approximately 100 iterations for the randomly initialized search to match the starting performance of the ProteinMPNN-MSD seeded population, highlighting the effectiveness of high-quality inverse folding priors. Furthermore, the ProteinMPNN-MSD initialized runs exhibit greater score variance, indicating that these priors may provide a more diverse set of starting candidates for exploring the multi-state landscape.

E Related Work

E.1 Protein Ensemble Generation Models

Recent advances in deep learning have revolutionized high-resolution protein structure prediction (Jumper et al., 2021; Baek et al., 2021) by leveraging evolutionary information from multiple sequence alignments (MSAs) as informative priors. To elicit conformational diversity from these models, several strategies have been developed to manipulate the input MSA, such as MSA subsampling (Del Alamo et al., 2022) and AFCluster (Wayment-Steele et al., 2024), the latter of which demonstrated that clustering MSAs can uncover alternative states in fold-switching proteins. Further refinements, including AFsample2 (Kalakoti & Wallner, 2025), improve conformational sampling through random MSA column masking to attenuate dominant co-evolutionary signals. Similarly, CFold (Bryant & Noé, 2024) employs explicit conformational data splits and integrates clustered MSAs to systematically explore structural landscapes.

Beyond MSA manipulation, generative and flow-based models have recently emerged to explicitly sample protein equilibrium ensembles. AlphaFlow (Jing et al., 2024) integrates flow matching with AlphaFold-based architectures to generate diverse structural states, while BioEmu (Lewis et al., 2025) utilizes a generative framework for the scalable emulation of conformational ensembles. Despite these advances, most existing approaches focus on single-chain monomers and often struggle to capture conformational transitions mediated by binding partners. In the context of multi-state design evaluation, DynamicMPNN (Abrudan et al., 2025) incorporates target-state structures as templates to guide conformation sampling. However, the use of explicit structural templates can bias the model toward reproducing specific conformations, potentially limiting the exploration of alternative functionally relevant states. Our work addresses these limitations by synergizing MSA clustering with the all-atom co-folding capabilities of AlphaFold3, enabling the resolution of structural ensembles across diverse biochemical contexts involving both protein and ligand binders.

E.2 Multi-State Protein Design Models

While conventional protein design focused on single-state protein structure using inverse-folding models such as ProteinMPNN (Dauparas et al., 2022), LigandMPNN (Dauparas et al., 2025) or ESM3 (Hayes et al., 2025), multi-state protein design aims to generate sequences that are compatible with multiple target conformations while maintaining appropriate energy gaps to avoid off-target states. Early approaches largely relied on physics-based energy functions and multi-state design protocols implemented within the Rosetta framework, which optimize sequences under multiple structural constraints through explicit energy landscape engineering. More recently, deep learning-based approaches have substantially improved design efficiency and scalability. ProteinMPNN-MSD (Dauparas et al., 2022) and ProteinGenerator (Lisanza et al., 2025) achieve multi-state compatibility by averaging sequence logits across multiple structural targets. DynamicMPNN (Abrudan et al., 2025) further advances this direction by training an inverse folding model conditioned on multiple conformational inputs. Caliby (Shuai et al., 2025) designs sequences by sampling from neural network–predicted, structure-conditioned Potts models, and generates sequences compatible with an ensemble of structures by averaging energies across multiple input structures. AF3-MSD builds on these foundations by utilizing AlphaFold3 as a structure-aware scoring proxy within an evolutionary framework, enabling the design of multi-state proteins with a focus on binder-mediated transitions.