RESTATEMPNN: STATE-RETENTIVE MODELING FOR MULTI-STATE PROTEIN INVERSE FOLDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Protein inverse folding is typically formulated for a single target structure, yet many functional proteins must remain compatible with multiple conformational states. Existing multi-state methods largely treat this setting as an information-fusion problem, combining state-conditioned predictions or aggregating state-indexed representations before sequence decoding. This formulation obscures structural changes between conformations and compresses state-specific variation before autoregressive sequence decisions are made. We introduce ReStateMPNN, a multi-state inverse-folding framework that jointly models cross-state structural relationships and preserves state-specific information throughout sequence generation. ReStateMPNN conditions cross-state communication on global conformational relations and residue-local contact rearrangements, while retaining state-specific residuals for sequence-context-dependent retrieval during decoding. Compared with ProteinMPNN$_{\mathrm{MSD}}$, ReStateMPNN improves sequence recovery from 38.0% to 45.0%, increases worst-state TM-score from 0.574 to 0.832, and reduces RMSD from 5.84 to 1.78 Å, with controlled ablations supporting complementary contributions from relation-conditioned communication and state-retentive decoding.

1 INTRODUCTION

Proteins often function through ensembles of interconverting conformations rather than a single static structure (Frauenfelder et al., 1991; Henzler-Wildman & Kern, 2007; Tokuriki & Tawfik, 2009; Nussinov et al., 2023). Transitions among these states underlie molecular recognition, catalysis, allostery, and other biological processes (Boehr et al., 2009; Henzler-Wildman et al., 2007; Motlagh et al., 2014). Metamorphic and fold-switching proteins provide pronounced examples in which one sequence adopts distinct native structures with different functional properties (Murzin, 2008; Porter & Looger, 2018; Dishman et al., 2021; Dishman & Volkman, 2023; Ruan et al., 2023). Recent advances in alternative-state prediction, conformational ensemble modeling, and dynamic protein engineering have expanded the methods available for studying and engineering alternative conformations (Janson et al., 2023; Bryant & Noé, 2024; Wayment-Steele et al., 2024; Praetorius et al., 2023; Pillai et al., 2024; Guo et al., 2025). Multi-state protein design therefore seeks a single sequence that remains compatible with multiple prescribed conformations rather than optimizing one structural state in isolation (Alberstein et al., 2022; Albanese et al., 2025).

Many inverse-folding models are developed and evaluated using individual target structures (Kuhlman & Bradley, 2019; Notin et al., 2024; Listov et al., 2024). Geometric graph and equivariant architectures encode backbone geometry through spatial neighborhoods, relative orientations, local coordinate frames, and scalar–vector features (Ingraham et al., 2019; Jing et al., 2021), forming the basis of models such as ESM-IF, ProteinMPNN, and PiFold (Hsu et al., 2022; Dauparas et al., 2022; Gao et al., 2023). Recent methods enrich this conditioning context with atom-level environments, ligands, side chains, and protein surfaces (Krapp et al., 2024; Dauparas et al., 2025; Li et al., 2026), while relation-aware encoders such as IEConv and GearNet explicitly parameterize geometric and typed structural interactions (Hermosilla et al., 2021; Zhang et al., 2023). These formulations characterize structural context within each conformation, whereas multi-state design also involves differences between the structural environments of corresponding residues across conformations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Learning-based multi-state methods primarily integrate conformational information through prediction-level fusion or representation-level aggregation (Praetorius et al., 2023; Lisanza et al., 2025; Joshi et al., 2025; Abrudan et al., 2026; Liu et al., 2025). ProteinMPNN$_{\mathrm{MSD}}$ and ProteinGenerator combine state-conditioned sequence predictions during generation (Praetorius et al., 2023; Lisanza et al., 2025). Representation-level approaches instead maintain state-indexed features before shared prediction: gRNAde introduces an explicit conformational axis with cross-state communication and permutation-invariant aggregation (Zaheer et al., 2017; Maron et al., 2020; Joshi et al., 2025); DynamicMPNN learns an ensemble-conditioned autoregressive sequence distribution from aligned conformations (Abrudan et al., 2026); and ABACUS-T$_{\mathrm{MSD}}$ performs inter-state representation updates followed by state pooling (Liu et al., 2025). In architectures that pool state-indexed representations, subsequent decoding uses aggregated features rather than direct access to each state representation. We investigate an alternative design that combines explicit relation descriptors with access to state-indexed residual features during decoding.

Main contributions. We introduce ReStateMPNN, a multi-state inverse-folding framework that uses explicit cross-state relational modeling and state-retentive autoregressive decoding. Our main contributions are: (i) Relation-Conditioned State Communication (RCSC). We replace latent-only cross-state interaction with explicit structural-relation conditioning, in which global conformational similarity and residue-local contact rewiring jointly determine which states communicate and how their messages are transformed. (ii) State-Retentive Autoregressive Decoding (SRAD). We avoid irreversible state collapse at ensemble aggregation by separating relation-updated representations into a shared consensus and centered state residuals, then retrieving the residuals from the autoregressive sequence context to provide state-specific corrections for each residue decision. (iii) Improved multi-state protein design. Compared with ProteinMPNN$_{\mathrm{MSD}}$, ReStateMPNN improves sequence recovery from 38.0% to 45.0%, increases TM-score from 0.689 to 0.907 and worst-state TM-score from 0.574 to 0.832, and reduces RMSD from 5.84 to 1.78 Å. Controlled ablations further isolate complementary gains from explicit relation conditioning and state-retentive decoding.

2 Related Work

Geometric Inverse Folding. Protein inverse-folding methods include geometric graph architectures and sequence models conditioned on protein structure. Early methods encode backbone geometry through spatial neighborhoods, pairwise distances, relative orientations, local coordinate frames, and invariant or equivariant features (Ingraham et al., 2019; Jing et al., 2021; Hsu et al., 2022; Dauparas et al., 2022; Gao et al., 2023). More recent approaches augment geometric encoding with pretrained protein representations, global structural context, iterative refinement, and alternative generative objectives, as exemplified by LM-Design, ProRefiner, KW-Design, CarbonDesign, Bridge-IF, and UniIF (Zheng et al., 2023; Zhou et al., 2023; Gao et al., 2024a; Ren et al., 2024; Zhu et al., 2024; Gao et al., 2024b). Structural conditioning has further expanded beyond backbone coordinates to atomic environments, ligands, side chains, protein surfaces, and multimodal sequence–structure representations (Krapp et al., 2024; Dauparas et al., 2025; Su et al., 2024; Li et al., 2024; 2026). Relation-aware encoders provide another line of development by explicitly modeling geometric or typed interactions within protein structures, including IEConv and GearNet (Hermosilla et al., 2021; Zhang et al., 2023).

Multi-State Protein Design. Multi-state protein design has been formulated as sequence optimization over multiple structural or energetic states (Ambroggio & Kuhlman, 2006; Leaver-Fay et al., 2011; Davey & Chica, 2012; Sauer et al., 2020; Löffler et al., 2017). Recent advances in alternative-state prediction, conformational ensemble modeling, and dynamic protein engineering have motivated learning-based formulations (Alberstein et al., 2022; Janson et al., 2023; Wayment-Steele et al., 2024; Bryant & Noé, 2024; Praetorius et al., 2023; Pillai et al., 2024; Guo et al., 2025). Prediction-level approaches, including ProteinMPNN$_{\mathrm{MSD}}$ and ProteinGenerator, combine state-conditioned sequence distributions during generation (Praetorius et al., 2023; Lisanza et al., 2025). Representation-level approaches instead integrate state-indexed structural features before shared sequence prediction. gRNAde maintains an explicit conformational axis with cross-state communication and permutation-invariant aggregation (Joshi et al., 2025); DynamicMPNN learns an ensemble-conditioned autoregressive distribution from aligned protein conformations (Abrudan et al., 2026); and ABACUS-T$_{\mathrm{MSD}}$ performs inter-state representation updates followed by state pooling (Liu et al., 2025).

3 BACKGROUND

Protein inverse folding. Protein inverse folding seeks an amino acid sequence compatible with a prescribed protein structure (Ingraham et al., 2019; Hsu et al., 2022; Dauparas et al., 2022; Gao et al., 2023). We represent a protein of length L by backbone coordinates $X = \{\mathbf{x}_{i,a} \in \mathbb{R}^3\}_{i=1,a=1}^{L,A}$ and sequence $Y = (y_1, \ldots, y_L)$, where $y_i \in \mathcal{A}$ and $|\mathcal{A}| = 20$. A structure-conditioned model parameterizes $p_\theta(Y \mid X)$. Autoregressive inverse-folding methods factorize this distribution according to a decoding order π as $p_\theta(Y \mid X, \pi) = \prod_{t=1}^{L} p_\theta(y_{\pi_t} \mid y_{\pi_{<t}}, X)$, so each residue prediction depends on the target structure and previously generated sequence context. This formulation underlies widely used structure-conditioned sequence models such as ESM-IF and ProteinMPNN (Hsu et al., 2022; Dauparas et al., 2022).

Geometric protein representations. Protein backbones are commonly represented as geometric graphs $G = (V, E)$ with residue nodes $V = \{v_i\}_{i=1}^{L}$ and edges determined by spatial or sequential neighborhoods (Ingraham et al., 2019; Jing et al., 2021; Dauparas et al., 2022; Zhang et al., 2023). Node and edge features encode structural information through distances, relative orientations, local coordinate frames, dihedral geometry, and invariant or equivariant scalar–vector features. Geometric Vector Perceptrons (GVPs), for example, maintain scalar channels that are invariant to rigid transformations together with vector channels that transform equivariantly under rotations (Jing et al., 2021). More generally, for a rigid transformation $T(X) = RX + \mathbf{t}$ with $R \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$, sequence prediction should satisfy $p_\theta(Y \mid T(X)) = p_\theta(Y \mid X)$. We denote the residue representation produced by a geometric encoder f_θ as $\mathbf{H} = f_\theta(X) = \{\mathbf{h}_i\}_{i=1}^{L}$. Relation-aware models such as IEConv and GearNet further augment this representation with explicitly parameterized geometric or typed interactions within a structure (Hermosilla et al., 2021; Zhang et al., 2023).

Multi-state inverse folding. Multi-state inverse folding extends the conditioning input from a single backbone to a conformational ensemble $\mathcal{X} = \{X^{(s)}\}_{s=1}^{S}$ and models a shared sequence distribution $p_\theta(Y \mid \mathcal{X})$ intended to generate sequences compatible with the prescribed states (Praetorius et al., 2023; Lisanza et al., 2025; Joshi et al., 2025; Abrudan et al., 2026; Liu et al., 2025). After establishing residue correspondence across structures, each conformation is encoded as $\mathbf{H}^{(s)} = \{\mathbf{h}_i^{(s)}\}_{i=1}^{L}$, with an observation mask $\mathbf{M}^{(s)} \in \{0,1\}^L$ for residues that are missing or unresolved in a particular state. Because conformations form an unordered set, the conditional distribution should be invariant to permutations of the state axis, i.e., $p_\theta(Y \mid X^{(1)}, \ldots, X^{(S)}) = p_\theta(Y \mid X^{(\sigma(1))}, \ldots, X^{(\sigma(S))})$ for any $\sigma \in \mathfrak{S}_S$.

Multi-state representation aggregation. Existing learning-based formulations combine conformational information either at the prediction level or within intermediate state representations (Praetorius et al., 2023; Lisanza et al., 2025; Joshi et al., 2025; Abrudan et al., 2026; Liu et al., 2025). Representation-level methods retain state-indexed features $\{\mathbf{h}_i^{(s)}\}_{s=1}^{S}$ during encoding and subsequently construct a shared representation for sequence prediction, commonly through a permutation-invariant operation such as $\bar{\mathbf{h}}_i = \mathrm{Pool}_s(\mathbf{h}_i^{(s)}, M_i^{(s)})$. Cross-state communication may be introduced before this aggregation, as in DSS-based multi-state encoding and inter-state representation updates (Maron et al., 2020; Joshi et al., 2025; Liu et al., 2025).

Notation. We use i, j for residue indices, s, t for conformational states, L for the number of aligned sequence positions, and S for the number of input states. $X^{(s)}$ denotes conformation s, $M_i^{(s)}$ its residue-observation mask, $\mathcal{S}_i = \{s : M_i^{(s)} = 1\}$ the valid state set at residue i, $\mathcal{X}$ the complete conformational ensemble, Y the designed sequence, and π an autoregressive decoding order. The shared geometric encoder produces invariant scalar features $\mathbf{H} = \{\mathbf{h}_{i,s}\}$ and equivariant vector features $\mathbf{V} = \{\mathbf{V}_{i,s}\}$. Pairwise conformational structure is summarized by relation descriptors $\mathbf{R} = \{\mathbf{r}_{i,s,t}\}$, and the corresponding state representations after cross-state communication are denoted by $\widetilde{\mathbf{H}} = \{\widetilde{\mathbf{h}}_{i,s}\}$. For each residue with $|\mathcal{S}_i| > 0$, we denote the masked state mean by $\boldsymbol{\mu}_i = |\mathcal{S}_i|^{-1} \sum_{s \in \mathcal{S}_i} \widetilde{\mathbf{h}}_{i,s}$ and the centered state deviation by $\boldsymbol{\delta}_{i,s} = \widetilde{\mathbf{h}}_{i,s} - \boldsymbol{\mu}_i$, with $\boldsymbol{\Delta}_i = \{\boldsymbol{\delta}_{i,s}\}_{s \in \mathcal{S}_i}$. The shared decoding pathway produces a sequence-context representation $\mathbf{q}_{\pi_t}$, and $\mathbf{c}_{\pi_t}$ denotes state-specific information retrieved from $\boldsymbol{\Delta}_{\pi_t}$ at decoding step t. The final amino-acid distribution is written as $p_\theta(y_{\pi_t} \mid Y_{\pi_{<t}}, \mathcal{X}, M)$. Bold lowercase symbols denote feature vectors, bold uppercase symbols denote collections or matrices of features, and calligraphic symbols denote sets.

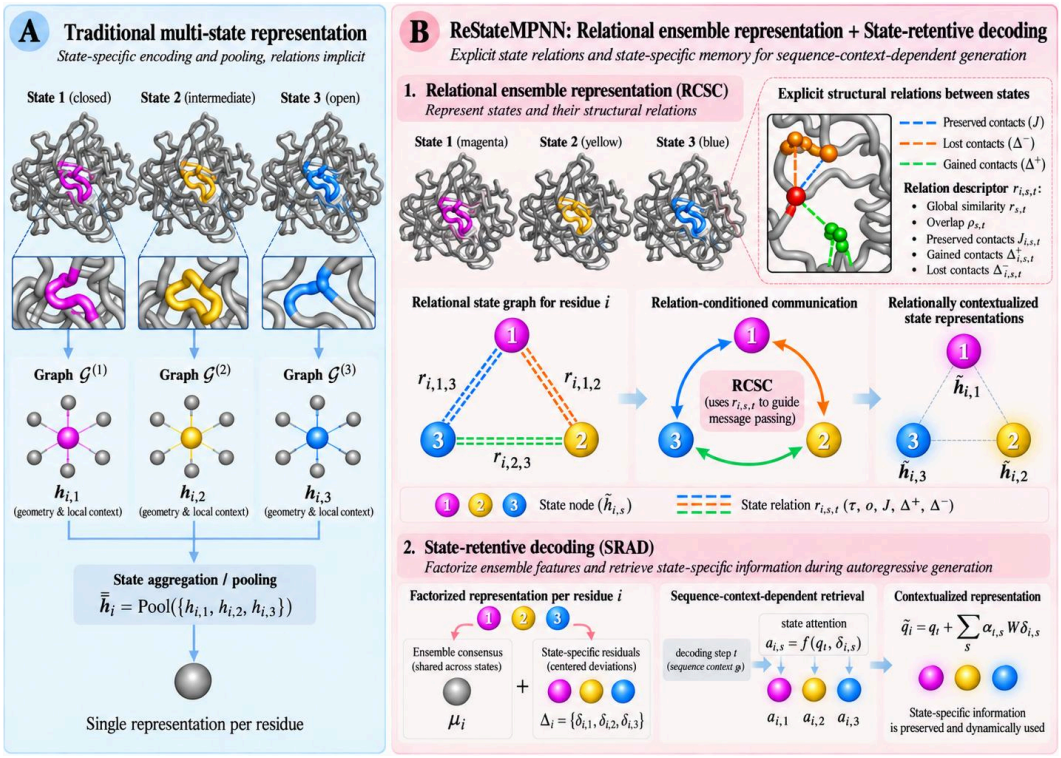

Figure 1: **Overview of ReStateMPNN.** (A) In pooling-based multi-state architectures, state-indexed features are aggregated into a shared representation, and individual state deviations may not be directly accessible during decoding. (B) ReStateMPNN employs relation-conditioned state communication (RCSC) to incorporate global and residue-local structural relations into cross-state interaction, and state-retentive autoregressive decoding (SRAD) to retrieve centered state residuals according to the evolving sequence context.

4 RESTATEMPNN

4.1 GEOMETRIC ENCODING

The conformational ensemble is represented in a shared Cartesian frame before geometric feature construction. For states s and t, let $\mathcal{O}_{s,t} = \{i : M_i^{(s)} = 1 \wedge M_i^{(t)} = 1\}$ denote the residues jointly observed in both structures. We denote by $d_{s,t}$ the C_α RMSD after optimal Kabsch superposition over $\mathcal{O}_{s,t}$ and by $o_{s,t} = |\mathcal{O}_{s,t}|/L$ the corresponding structural overlap. The ensemble reference is selected as an overlap-weighted structural medoid, after which every state is rigidly registered to that reference:

$$s^\star = \arg\min_s \frac{\sum_{t\neq s} o_{s,t} d_{s,t}}{\sum_{t\neq s} o_{s,t}},$$
$$(R_s^\star, \mathbf{t}_s^\star) = \arg\min_{R\in SO(3),\, \mathbf{t}\in\mathbb{R}^3} \sum_{i\in\mathcal{O}_{s,s^\star}} \left\| R\mathbf{x}_{i,C_\alpha}^{(s)} + \mathbf{t} - \mathbf{x}_{i,C_\alpha}^{(s^\star)} \right\|_2^2. \tag{1}$$

The aligned coordinates are $\hat{\mathbf{x}}_{i,a}^{(s)} = R_s^\star \mathbf{x}_{i,a}^{(s)} + \mathbf{t}_s^\star$, with $R_{s^\star}^\star = I$ and $\mathbf{t}_{s^\star}^\star = \mathbf{0}$. Medoid ties are resolved deterministically using structure-level identifiers. All node and edge geometric features are constructed from $\widehat{X}^{(s)}$.

Each aligned conformation is processed by the same geometric encoder, giving $(\mathbf{h}_{i,s}, \mathbf{V}_{i,s}) = E_\theta(\widehat{X}^{(s)}, M^{(s)})_i$, where $\mathbf{h}_{i,s} \in \mathbb{R}^{d_s}$ denotes invariant scalar features and $\mathbf{V}_{i,s} \in \mathbb{R}^{d_v\times 3}$ denotes equivariant vector features. Parameter sharing preserves permutation equivariance along the state axis. Multi-state interaction in ReStateMPNN is restricted to the scalar representation, while the

vector channels remain unchanged, $\widetilde{\mathbf{V}}_{i,s} = \mathbf{V}_{i,s}$. The aligned vector features are subsequently aggregated in the common ensemble frame as $\overline{\mathbf{V}}_i = |\mathcal{S}_i|^{-1} \sum_{s \in \mathcal{S}_i} \mathbf{V}_{i,s}$, with the same convention applied to vector-valued edge features. This retains the scalar–vector interface of the geometric backbone while isolating cross-state reasoning to rigid-motion-invariant representations.

4.2 Relation-Conditioned State Communication

Latent cross-state updates combine state-indexed representations without necessarily including separately computed structural-relation descriptors. RCSC augments this interaction with descriptors of both global conformational similarity and residue-local structural rearrangement.

Structural relation representation. For an ordered state pair (s, t), the global component contains the TM-score $\tau_{s,t}$ evaluated over the jointly observed residues together with their structural overlap $o_{s,t}$. The local component characterizes changes in the contact environment of residue i. Contacts are evaluated only over jointly observed residues, and short-range sequence neighbors on the same chain are excluded so that the descriptor emphasizes non-local packing changes. Let $\mathcal{C}_{i,s}^{(s,t)}$ and $\mathcal{C}_{i,t}^{(s,t)}$ denote the resulting neighborhoods in the two conformations. Their overlap is summarized by a Jaccard similarity $J_{i,s,t}$, while $\Delta_{i,s,t}^{+}$ and $\Delta_{i,s,t}^{-}$ measure normalized contact gain and loss from state s to state t. The relation descriptor is

$$\mathbf{r}_{i,s,t} = [\tau_{s,t},\, o_{s,t},\, J_{i,s,t},\, \Delta_{i,s,t}^{+},\, \Delta_{i,s,t}^{-}], \qquad \mathbf{e}_{i,s,t}^{R} = R_\rho(\mathbf{r}_{i,s,t}). \tag{2}$$

The global terms summarize structural similarity and residue overlap between conformations, whereas the local terms describe residue-specific neighborhood changes. The gain and loss components are directional, with $\Delta_{i,s,t}^{+} = \Delta_{i,t,s}^{-}$, so $\mathbf{r}_{i,s,t}$ can distinguish structurally asymmetric source–target interactions even when the global similarity is symmetric. Detailed contact definitions and normalization conventions are given in Appendix A.2.

Relation-conditioned state interaction. For target state s, RCSC aggregates information from the remaining observed states $t \in \mathcal{S}_i \setminus \{s\}$. A source state is characterized jointly by its latent compatibility with the target representation and by the explicit relation embedding $\mathbf{e}_{i,s,t}^{R}$. We denote the resulting relation-conditioned interaction operator by $\mathcal{F}_\rho$:

$$\widetilde{\mathbf{h}}_{i,s} = \mathbf{h}_{i,s} + \mathcal{F}_\rho\left(\mathbf{h}_{i,s}, \{\mathbf{h}_{i,t}, \mathbf{e}_{i,s,t}^{R}\}_{t \in \mathcal{S}_i \setminus \{s\}}\right). \tag{3}$$

Internally, $\mathcal{F}_\rho$ uses the relation embedding at two points. The first determines the allocation of information across source states, while the second modulates the content transferred from each source representation. Concretely, the state score combines latent query–key compatibility with a relation-dependent bias, and the source message undergoes relation-dependent multiplicative and additive modulation. These operations can be written compactly as

$$\begin{aligned} a_{i,s,t}^{\mathrm{rel}} &= \frac{Q(\mathbf{h}_{i,s})^\top K(\mathbf{h}_{i,t})}{\sqrt{d}} + b_\rho(\mathbf{e}_{i,s,t}^{R}), \\ \mathbf{m}_{i,s\leftarrow t} &= \left[1 + \tanh(W_\gamma \mathbf{e}_{i,s,t}^{R})\right] \odot W_V \mathbf{h}_{i,t} + W_\beta \mathbf{e}_{i,s,t}^{R}, \\ \widetilde{\mathbf{h}}_{i,s} &= \mathbf{h}_{i,s} + \eta_{\mathrm{rel}} \Gamma_\rho \left(\sum_{t \in \mathcal{S}_i \setminus \{s\}} \alpha_{i,s,t}^{\mathrm{rel}} \mathbf{m}_{i,s\leftarrow t} \right). \end{aligned} \tag{4}$$

The routing weights $\alpha_{i,s,t}^{\mathrm{rel}}$ are obtained by normalizing the state scores together with a zero-score, zero-value null candidate. We denote this normalization by $\mathrm{NullSoftmax}$. For scores $\{a_u\}_{u \in \mathcal{I}}$, the non-null mass is $\sum_{u \in \mathcal{I}} \alpha_u = 1 - \alpha_\varnothing$, so the same distribution specifies both the total amount of imported state information and its relative allocation across source conformations. With one available source state, the non-null weight reduces to a sigmoid of its interaction score rather than a one-element softmax. The explicit form and boundary cases of $\mathrm{NullSoftmax}$ are provided in Appendix A.3.

4.3 State-Retentive Autoregressive Decoding

Cross-state interaction produces an updated state-indexed representation $\{\widetilde{\mathbf{h}}_{i,s}\}_{s \in \mathcal{S}_i}$ for each residue. A shared decoder can use a permutation-invariant aggregate, but a pooling-only pathway

does not retain the individual state deviations as separately indexed inputs. SRAD retains these deviations beyond ensemble aggregation and conditions their retrieval on the evolving sequence context.

Consensus and residual representation. For residue i with $n_i = |\mathcal{S}_i| > 0$, the relation-updated states are represented by their ensemble consensus and centered state residuals:

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{s \in \mathcal{S}_i} \widetilde{\mathbf{h}}_{i,s}, \qquad \boldsymbol{\delta}_{i,s} = \widetilde{\mathbf{h}}_{i,s} - \boldsymbol{\mu}_i, \qquad \sum_{s \in \mathcal{S}_i} \boldsymbol{\delta}_{i,s} = \mathbf{0}. \tag{5}$$

The transformation is exactly invertible over the observed scalar state representations because $\widetilde{\mathbf{h}}_{i,s} = \boldsymbol{\mu}_i + \boldsymbol{\delta}_{i,s}$. Thus, state aggregation is separated from state retention: $\boldsymbol{\mu}_i$ provides the permutation-invariant representation consumed by the shared pathway, while $\boldsymbol{\Delta}_i = \{\boldsymbol{\delta}_{i,s}\}_{s \in \mathcal{S}_i}$ preserves deviations associated with individual conformations.

The consensus $\boldsymbol{\mu}_i$ is combined with the pooled vector representation $\overline{\mathbf{V}}_i$ and processed by the shared pooled geometric encoder. Denoting the resulting structural representation by $\mathbf{G} = \{\mathbf{g}_i\}_{i=1}^{L}$, the autoregressive decoder computes $\mathbf{q}_{\pi_t} = D_\phi(\mathbf{G}, Y_{\pi_{<t}})_{\pi_t}$. The decoder state therefore contains the ensemble-level structural representation together with the previously generated sequence context, while the state residuals remain outside the pooled pathway.

Context-conditioned residual retrieval. The residual-memory operator uses $\mathbf{q}_{\pi_t}$ to determine the relevance of each state deviation at the current decoding step. We write

$$\mathbf{c}^{\text{state}}_{\pi_t} = \mathcal{A}_\xi \left(\mathbf{q}_{\pi_t}, \boldsymbol{\Delta}_{\pi_t} \right), \qquad \widetilde{\mathbf{q}}_{\pi_t} = \mathbf{q}_{\pi_t} + \eta_{\text{mem}} \mathbf{c}^{\text{state}}_{\pi_t}, \tag{6}$$

where $\mathcal{A}_\xi$ is a shared state-retrieval operator. Its routing scores are computed from the compatibility between the decoder query and each centered residual, while the corresponding values are linear projections of the residual representations. Using the same null-augmented normalization as RCSC, the retrieval operation is

$$\begin{aligned} a^{\text{mem}}_{\pi_t,s} &= \frac{(W^m_Q \mathbf{q}_{\pi_t})^\top (W^m_K \boldsymbol{\delta}_{\pi_t,s})}{\sqrt{d_m}}, \boldsymbol{\beta}_{\pi_t,:} \quad = \text{NullSoftmax}\left(\{a^{\text{mem}}_{\pi_t,s}\}_{s \in \mathcal{S}_{\pi_t}} \right), \\ \mathbf{c}^{\text{state}}_{\pi_t} &= W_c \sum_{s \in \mathcal{S}_{\pi_t}} \beta_{\pi_t,s} W^m_V \boldsymbol{\delta}_{\pi_t,s}. \end{aligned} \tag{7}$$

The projections W^m_Q, W^m_K, W^m_V, and W_c are shared across states and contain no additive bias. No state-wise nonlinear transformation is applied to the residual value path before aggregation. These constraints preserve the centered structure of the memory representation through the value aggregation.

The residual formulation yields a direct distinction between shared and selective state information. If the retrieval distribution assigns equal weight to all observed states, then

$$\mathbf{c}^{\text{state}}_{\pi_t} = \beta_{\pi_t} W_c W^m_V \sum_{s \in \mathcal{S}_{\pi_t}} \boldsymbol{\delta}_{\pi_t,s} = \mathbf{0}. \tag{8}$$

Uniform state relevance therefore leaves the shared decoder representation unchanged. A nonzero correction arises only when the autoregressive context assigns non-uniform relevance to the centered deviations. The same construction gives zero residual correction when all state representations coincide and when only one state is observed. Formal derivations and boundary cases are reported in Appendix A.4.

4.4 Training Objective

The corrected decoder representation parameterizes the amino-acid distribution as $p_\theta(y_{\pi_t} \mid Y_{\pi_{<t}}, \mathcal{X}, M) = \text{softmax}(W_o \widetilde{\mathbf{q}}_{\pi_t} + \mathbf{b}_o)$. ReStateMPNN is trained by masked autoregressive maximum likelihood over valid sequence targets,

$$\mathcal{L}_{\text{seq}} = -\frac{1}{\sum_i m^{\text{loss}}_i} \sum_{t=1}^{L} m^{\text{loss}}_{\pi_t} \log p_\theta \left(y_{\pi_t} \mid Y_{\pi_{<t}}, \mathcal{X}, M \right). \tag{9}$$

The same target-validity policy is used across all training configurations. Positions without an observed structural state are excluded from state aggregation and handled according to this target mask. Architectural dimensions, relation cutoffs, frame-alignment conventions, label smoothing, optimization, and sampling settings are provided in Appendix B.3.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Datasets. We evaluate ReStateMPNN on the multi-state protein inverse folding benchmark introduced by DynamicMPNN (Abrudan et al., 2026). The benchmark contains 46,033 multi-conformational protein clusters constructed from experimentally resolved PDB and CoDNaS (Monzon et al., 2016) structures, where multiple conformations of the same sequence are treated as distinct structural states. The evaluation set consists of 96 proteins with experimentally observed conformational transitions and is used to evaluate sequence design across multiple structural states. Detailed dataset statistics and split procedures are provided in Appendix B.1.

Baselines. We compare ReStateMPNN with representative structure-conditioned inverse folding methods. ProteinMPNN (Dauparas et al., 2022) serves as the single-state autoregressive baseline, while multi-state methods include ProteinMPNN-$_{\text{MSD}}$, ProteinGenerator (Lisanza et al., 2025), DynamicMPNN (Abrudan et al., 2026), and ABACUS-T$_{\text{MSD}}$ (Liu et al., 2025), covering different strategies for integrating conformational information.

Evaluation Metrics. We evaluate both sequence fidelity and structural compatibility. Sequence-level performance is measured by sequence recovery and perplexity, while structural quality is assessed using pLDDT, RMSD, and TM-score across target conformations. We further report worst-state performance and multi-state compatibility metrics to evaluate whether designed sequences satisfy multiple structural constraints. Detailed metric definitions are provided in Appendix B.2.

Implementation Details. ReStateMPNN is trained end-to-end with the autoregressive negative log-likelihood objective using the training settings described in Appendix B.3 on four NVIDIA A800 GPUs.

Table 1: Performance on the multi-state protein inverse-folding benchmark. Results are reported for sequence design and structural recovery across target conformations. Best and second-best results are shown in **bold** and underline, respectively.

Model	Perplexity ↓	Recovery(%) ↑	pLDDT ↑	RMSD (Å) ↓	TM-score ↑	Worst-state TM ↑
ABACUS-T$_{\text{MSD}}$	4.41 (1.16)	42.1 (8.4)	81.14 (8.05)	2.78 (2.16)	0.845 (0.176)	0.742 (0.196)
ProteinGenerator	–	40.2 (9.5)	79.01 (8.74)	3.62 (2.95)	0.805 (0.198)	0.684 (0.223)
ProteinMPNN	5.46 (1.42)	36.7 (10.6)	71.86 (12.61)	6.67 (5.44)	0.651 (0.271)	0.536 (0.284)
ProteinMPNN$_{\text{MSD}}$	5.12 (1.34)	38.0 (11.0)	74.29 (11.37)	5.84 (4.56)	0.689 (0.244)	0.574 (0.263)
DynamicMPNN	4.28 (1.10)	42.7 (8.8)	82.26 (7.61)	2.29 (1.92)	0.872 (0.161)	0.781 (0.183)
ReStateMPNN w/o RCSC	4.19 (1.07)	43.7 (8.6)	82.91 (7.42)	2.11 (1.86)	0.883 (0.153)	0.795 (0.176)
ReStateMPNN w/o SRAD	4.15 (1.05)	44.0 (8.3)	83.28 (7.18)	2.03 (1.77)	0.890 (0.149)	0.806 (0.169)
ReStateMPNN	**3.96** (0.98)	**45.0** (8.0)	**84.37** (6.95)	**1.78** (1.55)	**0.907** (0.136)	**0.832** (0.154)

5.2 MULTI-STATE DESIGN PERFORMANCE

Table 1 summarizes performance on the multi-state inverse-folding benchmark. ReStateMPNN has the highest reported sequence recovery, pLDDT, and TM-scores and the lowest reported perplexity and RMSD among the methods in Table 1. Compared with DynamicMPNN, sequence recovery increases from 42.7% to 45.0%, while TM-score improves from 0.872 to 0.907 and RMSD decreases from 2.29 to 1.78 Å. ReStateMPNN also improves worst-state TM-score from 0.781 to 0.832, where worst-state TM-score is the minimum across the target conformations for each design. Table 2 further compares sequence recovery across training settings. ReStateMPNN reaches 45.0%, compared with 42.7% for Combined Pretraining + Multi Finetuning and 38.0% for ProteinMPNN$_{\text{MSD}}$.

Table 2: Sequence recovery on the multi-state design benchmark.

Model Variant	Recovery (%) ↑
ReStateMPNN	**45.0** (8.0)
Combined Pre. + Multi FT	42.7 (8.8)
Single Pre. + Multi FT	42.1 (8.3)
Combined Training	41.0 (8.5)
ProteinMPNN$_{\text{MSD}}$	38.0 (11.0)
Single Training	27.1 (9.4)

5.3 TEMPLATE-CONDITIONED STRUCTURAL SELF-CONSISTENCY

We further evaluate template-conditioned structural self-consistency using AF3, reporting both raw and decoy-normalized metrics (Table 3).

Structural recovery. Among the methods in Table 3, ReStateMPNN has the highest mean pLDDT and TM-score and the lowest mean RMSD. Compared with DynamicMPNN, pLDDT increases from 82.26 to 84.37 and TM-score from 0.872 to 0.907, while RMSD decreases from 2.29 to 1.78 Å.

Decoy-normalized AF3 metrics. Relative to DynamicMPNN, ReStateMPNN has a decoy-normalized pLDDT ratio of 1.538 versus 1.470, a TM-score ratio of 7.081 versus 6.630, and an RMSD ratio of 0.096 versus 0.125.

Table 3: Target-template AF3 structural self-consistency on the multi-state protein design benchmark. Raw metrics report agreement between target-template AF3 predictions and the corresponding target structures; decoy-normalized metrics compare these values with predictions obtained using structurally dissimilar decoy templates.

	Raw Metrics			Decoy-Normalized Metrics		
Model	**pLDDT ↑**	**RMSD (Å) ↓**	**TM-score ↑**	**pLDDT ↑**	**RMSD ↓**	**TM-score ↑**
ABACUS-T_{MSD}	81.14 (8.05)	2.78 (2.16)	0.845 (0.176)	1.455 (0.439)	0.145 (0.143)	6.604 (1.703)
ProteinGenerator	79.01 (8.74)	3.62 (2.95)	0.805 (0.198)	1.438 (0.447)	0.181 (0.163)	6.574 (1.921)
ProteinMPNN	71.86 (12.61)	6.67 (5.44)	0.651 (0.271)	1.372 (0.431)	0.318 (0.251)	5.912 (2.493)
ProteinMPNN$_{MSD}$	74.29 (11.37)	5.84 (4.56)	0.689 (0.244)	1.416 (0.389)	0.275 (0.219)	6.533 (2.325)
DynamicMPNN	82.26 (7.61)	2.29 (1.92)	0.872 (0.161)	1.470 (0.466)	0.125 (0.129)	6.630 (1.506)
ReStateMPNN w/o RCSC	82.91 (7.42)	2.11 (1.86)	0.883 (0.153)	1.492 (0.449)	0.116 (0.122)	6.782 (1.524)
ReStateMPNN w/o SRAD	83.28 (7.18)	2.03 (1.77)	0.890 (0.149)	1.506 (0.441)	0.109 (0.118)	6.903 (1.468)
ReStateMPNN	**84.37** (6.95)	**1.78** (1.55)	**0.907** (0.136)	**1.538** (0.432)	**0.096** (0.107)	**7.081** (1.412)

5.4 ABLATION STUDIES

Relation-conditioned state communication. Table 4 reports sequence recovery and decoy-normalized metrics for five RCSC configurations. Sequence recovery increases from 43.7% without cross-state communication to 43.9% with latent-state communication, and further to 44.2% and 44.6% using global and local relations, respectively. The configuration using both relation types records 45.0% recovery and a normalized TM-score ratio of 7.081; these values are higher than those of the corresponding single-relation variants in this ablation.

Table 4: Ablation of Relation-Conditioned State Communication (RCSC).

Variant	**Interaction**	**Global relation**	**Local relation**	**Recovery(%) ↑**	**pLDDT ↑**	**RMSD ↓**	**TM-score ↑**
No cross-state commu.	×	×	×	43.7 (8.6)	1.492 (0.449)	0.116 (0.122)	6.782 (1.524)
Latent-state commun.	✓	×	×	43.9 (8.5)	1.503 (0.445)	0.111 (0.117)	6.861 (1.497)
Global relation only	✓	✓	×	44.2 (8.4)	1.512 (0.438)	0.105 (0.113)	6.944 (1.472)
Local relation only	✓	×	✓	44.6 (8.7)	1.532 (0.441)	0.101 (0.109)	7.012 (1.446)
ReStateMPNN	✓	✓	✓	**45.0** (8.0)	**1.538** (0.432)	**0.096** (0.107)	**7.081** (1.412)

State-retentive autoregressive decoding. Table 5 reports the results for four SRAD configurations. Pooling alone achieves 44.0% recovery and a normalized TM-score of 6.903. Retaining state residuals improves these values to 44.3% and 6.958, while dynamic retrieval further increases them to 44.6% and 7.019. The full sequence-context-conditioned formulation reaches 45.0% recovery and 7.081 normalized TM-score, higher than the corresponding pooling-only values in this ablation.

Table 5: Ablation of State-Retentive Autoregressive Decoding (SRAD).

Variant	**State retained**	**Dynamic retrieval**	**Conditioned**	**Recovery(%) ↑**	**pLDDT ↑**	**RMSD ↓**	**TM-score ↑**
Pooling only	×	×	×	44.0 (8.3)	1.506 (0.441)	0.109 (0.118)	6.903 (1.468)
Residual retention	✓	×	×	44.3 (8.1)	1.515 (0.436)	0.105 (0.114)	6.958 (1.453)
Retrieval	✓	✓	×	44.6 (8.4)	1.526 (0.439)	0.110 (0.110)	7.019 (1.438)
ReStateMPNN	✓	✓	✓	**45.0** (8.0)	**1.538** (0.432)	**0.096** (0.107)	**7.081** (1.412)

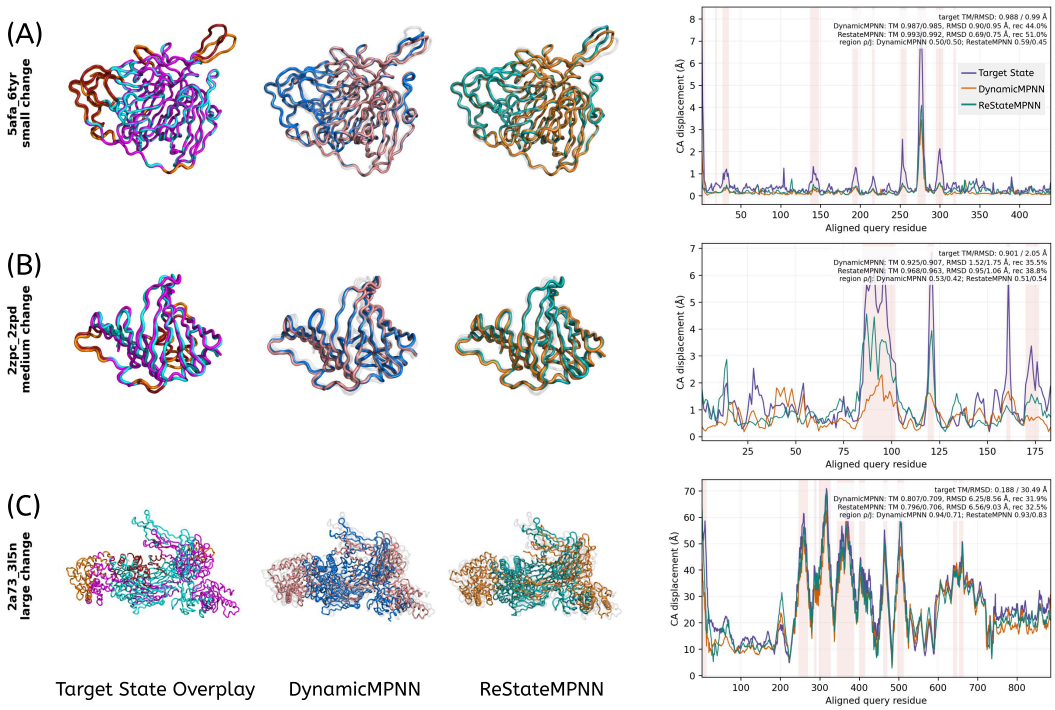

Figure 2: **Structural case studies across conformational changes.** Target conformations and AF3 structures from the designed sequences are shown together with residue-wise Cα displacement profiles. Red shading marks the top 20% most displaced target residues. (A) 5afa_6tyr, small local change; (B) 2zpc_2zpd, medium rearrangement; (C) 2a73_3l5n, large-scale transition.

5.5 Structural Case Studies

We examine three selected protein pairs spanning small, medium, and large conformational changes (Fig. 2). For both DynamicMPNN and ReStateMPNN, we use `sample_000`, with AF3 seed 0 and a single diffusion sample; no best-of-N selection is applied. We compare the target conformations with the resulting AF3 structures and their residue-wise Cα displacement profiles.

For the small-change 5afa_6tyr pair (Fig. 2A), the prediction of the target-template AF3 of both designed sequences closely matches the respective target structures. ReStateMPNN improves the matched-state TM-score from 0.987/0.985 to 0.993/0.992 and reduces RMSD from 0.90/0.95 to 0.69/0.75 Å. For the medium 2zpc_2zpd transition (Fig. 2B), the difference is larger: ReStateMPNN improves TM-score from 0.925/0.907 to 0.968/0.963 and reduces RMSD from 1.52/1.75 to 0.95/1.06 Å. The overlap of high-displacement residues also increases from $J = 0.42$ to $J = 0.54$. The large-transition 2a73_3l5n pair (Fig. 2C) remains challenging for both methods. The target states differ substantially (TM-score $= 0.188$, RMSD $= 30.49$ Å), while target-template AF3 predictions for both methods show similar displacement-profile correlations ($\rho = 0.94$ for DynamicMPNN and 0.93 for ReStateMPNN). DynamicMPNN gives a slightly better global TM-score and RMSD, whereas ReStateMPNN achieves a larger overlap with the most displaced residues ($J = 0.83$ vs. 0.71).

6 Conclusion

We present ReStateMPNN, a multi-state inverse-folding framework that combines relation-conditioned state communication with state-retentive autoregressive decoding. By modeling explicit cross-conformation relations while preserving state-specific information beyond ensemble aggregation, we have improved sequence recovery and conformational refoldability, including worst-state compatibility. Large-scale transitions remain challenging, motivating richer ensemble modeling and experimental validation.

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 Knowles. Multi-state protein sequence design with dynamicMPNN. In *ICLR*, 2026.

Katherine I. Albanese, Sophie Barbe, Shunsuke Tagami, Derek N. Woolfson, and Thomas Schiex. Computational protein design. *Nature Reviews Methods Primers*, 5(1):13, 2025.

Robert G. Alberstein, Amy B. Guo, and Tanja Kortemme. Design principles of protein switches. *Current Opinion in Structural Biology*, 72:71–78, 2022.

Xavier I. Ambroggio and Brian Kuhlman. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *Journal of the American Chemical Society*, 128(4):1154–1161, 2006.

David D. Boehr, Ruth Nussinov, and Peter E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, 2009.

Patrick Bryant and Frank Noé. Structure prediction of alternative protein conformations. *Nature Communications*, 15(1):7328, 2024. ISSN 2041-1723.

J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. 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*, 22(4):717–723, 2025. ISSN 1548-7105.

James A. Davey and Roberto A. Chica. Multistate approaches in computational protein design. *Protein Science*, 21(9):1241–1252, 2012.

Acacia F. Dishman and Brian F. Volkman. Metamorphic protein folding as evolutionary adaptation. *Trends in Biochemical Sciences*, 48(8):665–672, 2023.

Acacia F. Dishman, Robert C. Tyler, Jamie C. Fox, Andrew B. Kleist, Kenneth E. Prehoda, M. Madan Babu, Francis C. Peterson, and Brian F. Volkman. Evolution of fold switching in a metamorphic protein. *Science*, 371(6524):86–90, 2021.

Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

Zhangyang Gao, Cheng Tan, and Stan Z. Li. Pifold: Toward effective and efficient protein inverse folding. In *ICLR*, 2023.

Zhangyang Gao, Cheng Tan, Xingran Chen, Yijie Zhang, Jun Xia, Siyuan Li, and Stan Z. Li. Kw-design: Pushing the limit of protein design via knowledge refinement. In *ICLR*, 2024a.

Zhangyang Gao, Jue Wang, Cheng Tan, Lirong Wu, Yufei Huang, Siyuan Li, Zhirui Ye, and Stan Z. Li. Uniif: Unified molecule inverse folding. In *NeurIPS*, 2024b.

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

Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450 (7172):964–972, 2007.

Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450 (7171):913–916, 2007.

Pedro Hermosilla, Marco Schäfer, Matej Lang, Gloria Fackelmann, Pere-Pau Vázquez, Barbora Kozlíková, Michael Krone, Tobias Ritschel, and Timo Ropinski. Intrinsic-extrinsic convolution and pooling for learning on 3d protein structures. In *ICLR*, 2021.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *ICML*, volume 162, pp. 8946–8970, 2022.

John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. *NeurIPS*, 32:15820–15831, 2019.

Giacomo Janson, Gilberto Valdes-Garcia, Lim Heo, and Michael Feig. Direct generation of protein conformational ensembles via machine learning. *Nature Communications*, 14(1):774, 2023.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron O. Dror. Learning from protein structure with geometric vector perceptrons. In *ICLR*, 2021.

Chaitanya K. Joshi, Arian Rokkum Jamasb, Ramon Viñas Torné, Charles Harris, Simon V Mathis, Alex Morehead, Rishabh Anand, and Pietro Lio. gRNAde: Geometric deep learning for 3d RNA inverse design. In *ICLR*, 2025.

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.

Lucien F. Krapp, Fernando A. Meireles, Luciano A. Abriata, Jean Devillard, Sarah Vacle, Maria J. Marcaida, and Matteo Dal Peraro. Context-aware geometric deep learning for protein sequence design. *Nature Communications*, 15(1):6273, 2024.

Brian Kuhlman and Philip Bradley. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 20(11):681–697, 2019.

Andrew Leaver-Fay, Ron Jacak, P. Benjamin Stranges, and Brian Kuhlman. A generic program for multistate protein design. *PLoS ONE*, 6(7):e20937, 2011.

Darong Li, Lian Shen, Meijia Song, Deyi Li, Juan Liu, and Xiangrong Liu. Surffold: a unified model for protein inverse folding by integrating surface and structural information. *Bioinformatics*, 42 (1):btaf666, 2026. ISSN 1367-4811.

Mingchen Li, Yang Tan, Xinzhu Ma, Bozitao Zhong, Huiqun Yu, Ziyi Zhou, Wanli Ouyang, Bingxin Zhou, Pan Tan, and Liang Hong. Prosst: Protein language modeling with quantized structure and disentangled attention. In *NeurIPS*, 2024.

Sidney Lyayuga Lisanza, Jacob Merle Gershon, Samuel W. K. Tipps, Jeremiah Nelson Sims, and et al. Arnoldt, Lucas. Multistate and functional protein design using rosettafold sequence space diffusion. *Nature Biotechnology*, 43(8):1288–1298, 08 2025.

Dina Listov, Casper A. Goverde, Bruno E. Correia, and Sarel Jacob Fleishman. Opportunities and challenges in design and optimization of protein function. *Nature Reviews Molecular Cell Biology*, 25(8):639–653, 2024.

Yufeng Liu, Rui Wu, Xinyu Wang, Sheng Wang, Linghui Chen, Fudong Li, Quan Chen, and Haiyan Liu. Enhancing functional proteins through multimodal inverse folding with abacus-t. *Nature Communications*, 16:10177, 11 2025.

Patrick Löffler, Samuel Schmitz, Enrico Hupfeld, Reinhard Sterner, and Rainer Merkl. Rosetta:msf: a modular framework for multi-state computational protein design. *PLOS Computational Biology*, 13(6):e1005600, 2017.

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lddt: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On learning sets of symmetric elements. In *ICML*, volume 119, pp. 6734–6744, 2020.

Alexander M. Monzon, Christian O. Rohr, Maria Silvina Fornasari, and Gustavo Parisi. Codnas 2.0: a comprehensive database of protein conformational diversity in the native state. *Database*, 2016: baw038, 2016.

Hesam N. Motlagh, James O. Wrabl, Jing Li, and Vincent J. Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, 2014.

Alexey G. Murzin. Metamorphic proteins. *Science*, 320(5884):1725–1726, 2008.

Pascal Notin, Nathan Rollins, Yarin Gal, Chris Sander, and Debora Marks. Machine learning for functional protein design. *Nature Biotechnology*, 42(2):216–228, 2024.

Ruth Nussinov, Yonglan Liu, Wengang Zhang, and Hyunbum Jang. Protein conformational ensembles in function: roles and mechanisms. *RSC Chemical Biology*, 4(11):850–864, 2023.

Arvind Pillai, Abbas Idris, Annika Philomin, Connor Weidle, Rebecca Skotheim, Philip J. Y. Leung, Adam Broerman, Cullen Demakis, Andrew J. Borst, Florian Praetorius, and David Baker. De novo design of allosterically switchable protein assemblies. *Nature*, 632(8026):911–920, Aug 2024. ISSN 1476-4687.

Lauren L. Porter and Loren L. Looger. Extant fold-switching proteins are widespread. *Proceedings of the National Academy of Sciences*, 115(23):5968–5973, 2018.

Florian Praetorius, Philip J. Y. Leung, Maxx H. Tessmer, Adam Broerman, Cullen Demakis, Acacia F. Dishman, Arvind Pillai, Abbas Idris, David Juergens, Justas Dauparas, Xinting Li, Paul M. Levine, Mila Lamb, Ryanne K. Ballard, Stacey R. Gerben, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Brian F. Volkman, Jeff Nivala, Stefan Stoll, and David Baker. Design of stimulus-responsive two-state hinge proteins. *Science*, 381(6659):754–760, 2023.

Milong Ren, Chungong Yu, Dongbo Bu, and Haicang Zhang. Accurate and robust protein sequence design with CarbonDesign. *Nature Machine Intelligence*, 6(5):536–547, 2024.

James P. Roney and Sergey Ovchinnikov. State-of-the-art estimation of protein model accuracy using alphafold. *Physical Review Letters*, 129:238101, 2022.

Biao Ruan, Yanan He, Yingwei Chen, Eun Jung Choi, Yihong Chen, Dana Motabar, Tsega Solomon, Richard Simmerman, Thomas Kauffman, D. Travis Gallagher, John Orban, and Philip N. Bryan. Design and characterization of a protein fold switching network. *Nature Communications*, 14(1): 431, 2023.

Marion F. Sauer, Alexander M. Sevy, James E. Crowe, and Jens Meiler. Multi-state design of flexible proteins predicts sequences optimal for conformational change. *PLOS Computational Biology*, 16(2):e1007339, 2020.

Jin Su, Chenchen Han, Yuyang Zhou, Junjie Shan, Xibin Zhou, and Fajie Yuan. Saprot: Protein language modeling with structure-aware vocabulary. In *ICLR*, 2024.

Nobuhiko Tokuriki and Dan S. Tawfik. Protein dynamism and evolvability. *Science*, 324(5924): 203–207, 2009.

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

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander J Smola. Deep Sets. In *NIPS*, pp. 3391–3401, 2017.

Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

Yang Zhang and Jeffrey Skolnick. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005.

Zuobai Zhang, Minghao Xu, Arian Rokkum Jamasb, Vijil Chenthamarakshan, Aurélie C. Lozano, Payel Das, and Jian Tang. Protein representation learning by geometric structure pretraining. In *ICLR*, 2023.

Zaixiang Zheng, Yifan Deng, Dongyu Xue, Yi Zhou, Fei Ye, and Quanquan Gu. Structure-informed language models are protein designers. In *ICML*, volume 202, pp. 42317–42338, 2023.

Xinyi Zhou, Guangyong Chen, Junjie Ye, Ercheng Wang, Jun Zhang, Cong Mao, Zhanwei Li, Jianye Hao, Xingxu Huang, Jin Tang, and Pheng Ann Heng. Prorefiner: an entropy-based refining strategy for inverse protein folding with global graph attention. *Nature Communications*, 14(1): 7434, 2023.

Yiheng Zhu, Jialu Wu, Qiuyi Li, Jiahuan Yan, Mingze Yin, Wei Wu, Mingyang Li, Jieping Ye, Zheng Wang, and Jian Wu. Bridge-if: Learning inverse protein folding with markov bridges. In *NeurIPS*, 2024.

A METHOD DETAILS

A.1 COMMON-FRAME GEOMETRIC REPRESENTATION

Section 4.1 represents all conformational states in a common Cartesian frame before geometric encoding. Here we give the complete construction and characterize its transformation behavior under independently chosen coordinate frames.

For states s and t, let $\mathcal{O}_{s,t} = \{i : M_i^{(s)} = 1 \wedge M_i^{(t)} = 1\}$ denote the residues observed in both conformations. Their structural discrepancy is defined by the optimal rigid superposition of the corresponding C_α coordinates,

$$d_{s,t}^2 = \min_{R\in SO(3),\,\mathbf{t}\in\mathbb{R}^3} \frac{1}{|\mathcal{O}_{s,t}|} \sum_{i\in\mathcal{O}_{s,t}} \left\| R\mathbf{x}_{i,C_\alpha}^{(s)} + \mathbf{t} - \mathbf{x}_{i,C_\alpha}^{(t)} \right\|_2^2, \qquad o_{s,t} = \frac{|\mathcal{O}_{s,t}|}{L},$$
$$s^\star = \arg\min_s \frac{\sum_{t\neq s} o_{s,t} d_{s,t}}{\sum_{t\neq s} o_{s,t}}. \tag{10}$$

The overlap term $o_{s,t}$ weights each pairwise comparison by the fraction of residues supporting it. The resulting $s^\star$ is therefore an overlap-weighted structural medoid of the ensemble. When multiple states attain the same medoid objective, the tie is resolved by a deterministic structure-level identifier that is independent of the state tensor ordering.

Each state is then registered to $s^\star$ over their common observed residues. Denoting the optimal transformation by $T_s^\star = (R_s^\star, \mathbf{t}_s^\star)$,

$$(R_s^\star, \mathbf{t}_s^\star) = \arg\min_{R\in SO(3),\,\mathbf{t}\in\mathbb{R}^3} \sum_{i\in\mathcal{O}_{s,s^\star}} \left\| R\mathbf{x}_{i,C_\alpha}^{(s)} + \mathbf{t} - \mathbf{x}_{i,C_\alpha}^{(s^\star)} \right\|_2^2,$$
$$\widehat{\mathbf{x}}_{i,a}^{(s)} = R_s^\star \mathbf{x}_{i,a}^{(s)} + \mathbf{t}_s^\star, \tag{11}$$

with $T_{s^\star}^\star = (I, \mathbf{0})$. The same transformation is applied to every observed atom in state s, so all node and edge geometric features are constructed only after the state has been mapped into the ensemble frame.

We next consider arbitrary rigid coordinate changes applied independently to the input states. Let $G_s(\mathbf{x}) = A_s\mathbf{x} + \mathbf{b}_s$, where $A_s \in SO(3)$, and define $X^{(s)\prime} = G_s X^{(s)}$. For any candidate alignment $(R, \mathbf{t})$ between the original states, define

$$R' = A_t R A_s^\top, \qquad \mathbf{t}' = A_t\mathbf{t} + \mathbf{b}_t - R'\mathbf{b}_s.$$

The corresponding residual in the transformed coordinates satisfies

$$R'G_s(\mathbf{x}_{i,C_\alpha}^{(s)}) + \mathbf{t}' - G_t(\mathbf{x}_{i,C_\alpha}^{(t)}) = A_t\left(R\mathbf{x}_{i,C_\alpha}^{(s)} + \mathbf{t} - \mathbf{x}_{i,C_\alpha}^{(t)}\right),$$
$$\left\| R'G_s(\mathbf{x}_{i,C_\alpha}^{(s)}) + \mathbf{t}' - G_t(\mathbf{x}_{i,C_\alpha}^{(t)}) \right\|_2 = \left\| R\mathbf{x}_{i,C_\alpha}^{(s)} + \mathbf{t} - \mathbf{x}_{i,C_\alpha}^{(t)} \right\|_2. \tag{12}$$

Because the mapping $(R, \mathbf{t}) \mapsto (R', \mathbf{t}')$ is bijective over rigid transformations, minimizing both sides yields $d'_{s,t} = d_{s,t}$. The observation overlap is also unchanged, $o'_{s,t} = o_{s,t}$, and hence the medoid objective in Eq. 10 is invariant to independent rigid transformations of the states. The selected reference therefore remains $s^\star$, subject only to the deterministic tie convention above.

It remains to characterize the registered coordinates themselves. If $T_s^\star$ maps the original state s into the reference frame, then an optimal registration for the independently transformed inputs is $T_s^{\star\prime} = G_{s^\star} \circ T_s^\star \circ G_s^{-1}$.

$$\widehat{X}^{(s)\prime} = T_s^{\star\prime} X^{(s)\prime} = G_{s^\star} T_s^\star G_s^{-1} G_s X^{(s)} = G_{s^\star}\widehat{X}^{(s)}. \tag{13}$$

Thus, independent rigid transformations of the raw conformations induce only a single common rigid transformation after registration.

Let $A_{s^\star}$ denote the rotational component of this common transformation. For an equivariant geometric encoder,

$$\mathbf{h}'_{i,s} = \mathbf{h}_{i,s}, \qquad \mathbf{V}'_{i,s} = A_{s^\star}\mathbf{V}_{i,s}.$$

The state-pooled vector representation therefore satisfies.

$$\overline{\mathbf{V}}'_i = \frac{1}{|\mathcal{S}_i|} \sum_{s \in \mathcal{S}_i} A_{s^\star} \mathbf{V}_{i,s} = A_{s^\star} \overline{\mathbf{V}}_i.$$

A.2 Structural Relation Formulation

For an ordered pair of conformational states (s, t), structural relations are defined over the common residue support $\mathcal{O}_{s,t} = \{i : M_i^{(s)} = 1 \wedge M_i^{(t)} = 1\}$. The global component describes the similarity and coverage of the state pair, while the local component characterizes the change in the structural neighborhood of each residue.

The global relation is represented by

$$\mathbf{r}^G_{s,t} = [\tau_{s,t},\, o_{s,t}], \qquad o_{s,t} = \frac{|\mathcal{O}_{s,t}|}{L}, \tag{14}$$

where $\tau_{s,t}$ is the TM-score evaluated on the coordinates indexed by $\mathcal{O}_{s,t}$. Both components are symmetric with respect to exchanging the two states, $\tau_{s,t} = \tau_{t,s}$ and $o_{s,t} = o_{t,s}$.

For residue $i \in \mathcal{O}_{s,t}$, the local relation is defined from its non-local contact neighborhood in each conformation. Let $c(i)$ denote the chain identity of residue i and $p(i)$ its within-chain sequence index. The contact set of residue i in state s, restricted to residues jointly observed in the state pair, is

$$\mathcal{C}^{(s,t)}_{i,s} = \left\{ j \in \mathcal{O}_{s,t} \setminus \{i\} \;\middle|\; d^{(s)}_{ij} < d_0,\; c(i) \neq c(j) \;\vee\; |p(i) - p(j)| > k_{\text{seq}} \right\}, \tag{15}$$

with $\mathcal{C}^{(s,t)}_{i,t}$ defined analogously. The sequence-separation condition removes short-range same-chain contacts while retaining inter-chain interactions. The residue-pair distance convention and the values of d_0 and k_{seq} are specified in Appendix B.3.

The transition from state s to state t decomposes the two contact sets into preserved, gained, and lost contacts,

$$\begin{aligned} \mathcal{P}_{i,s,t} &= \mathcal{C}^{(s,t)}_{i,s} \cap \mathcal{C}^{(s,t)}_{i,t}, \\ \mathcal{G}_{i,s,t} &= \mathcal{C}^{(s,t)}_{i,t} \setminus \mathcal{C}^{(s,t)}_{i,s}, \qquad \mathcal{U}_{i,s,t} = \mathcal{P}_{i,s,t} \,\dot\cup\, \mathcal{G}_{i,s,t} \,\dot\cup\, \mathcal{L}_{i,s,t}, \\ \mathcal{L}_{i,s,t} &= \mathcal{C}^{(s,t)}_{i,s} \setminus \mathcal{C}^{(s,t)}_{i,t}, \end{aligned} \tag{16}$$

where $\mathcal{U}_{i,s,t} = \mathcal{C}^{(s,t)}_{i,s} \cup \mathcal{C}^{(s,t)}_{i,t}$ and $\dot\cup$ denotes a disjoint union. The three sets therefore partition the complete contact support involved in the conformational transition.

Normalizing their cardinalities by $|\mathcal{U}_{i,s,t}|$ gives

$$J_{i,s,t} = \frac{|\mathcal{P}_{i,s,t}|}{|\mathcal{U}_{i,s,t}|}, \qquad \Delta^+_{i,s,t} = \frac{|\mathcal{G}_{i,s,t}|}{|\mathcal{U}_{i,s,t}|}, \qquad \Delta^-_{i,s,t} = \frac{|\mathcal{L}_{i,s,t}|}{|\mathcal{U}_{i,s,t}|}. \tag{17}$$

For $|\mathcal{U}_{i,s,t}| = 0$, the convention $J_{i,s,t} = 1$ and $\Delta^+_{i,s,t} = \Delta^-_{i,s,t} = 0$ assigns identical empty neighborhoods maximal preservation and zero rearrangement. In either case,

$$J_{i,s,t} + \Delta^+_{i,s,t} + \Delta^-_{i,s,t} = 1. \tag{18}$$

The local descriptor therefore constitutes a normalized decomposition of the residue-level contact transition into preserved, gained, and lost structural interactions.

Reversing the ordered state pair preserves the overlap term while exchanging gain and loss:

$$J_{i,t,s} = J_{i,s,t}, \qquad \Delta^+_{i,t,s} = \Delta^-_{i,s,t}, \qquad \Delta^-_{i,t,s} = \Delta^+_{i,s,t}. \tag{19}$$

The symmetric component $J_{i,s,t}$ measures contact conservation, whereas $(\Delta^+_{i,s,t}, \Delta^-_{i,s,t})$ retains the direction of the local structural transition.

Combining the state-level and residue-level terms yields the relation descriptor used by RCSC,

$$\mathbf{r}_{i,s,t} = [\tau_{s,t},\, o_{s,t},\, J_{i,s,t},\, \Delta^+_{i,s,t},\, \Delta^-_{i,s,t}], \qquad \mathbf{r}_{i,t,s} = [\tau_{s,t},\, o_{s,t},\, J_{i,s,t},\, \Delta^-_{i,s,t},\, \Delta^+_{i,s,t}]. \tag{20}$$

The first three components are invariant to reversing the state pair, whereas the final two encode the ordered source–target change. Because $\tau_{s,t}$ is evaluated after structural superposition, $o_{s,t}$ depends only on observation masks, and the local terms depend only on inter-residue distances and contact-set operations, $\mathbf{r}_{i,s,t}$ is invariant to rigid transformations of either conformation.

A.3 Relation-Conditioned Cross-State Update

For residue i and target state s, let $\mathcal{I}_{i,s} = \mathcal{S}_i \setminus \{s\}$ denote the set of valid source states. The structural descriptor $\mathbf{r}_{i,s,t}$ defined in Appendix A.2 is embedded as $\mathbf{e}^R_{i,s,t} = R_\rho(\mathbf{r}_{i,s,t})$. The RCSC operator in Eq. 3 conditions cross-state interaction on $\mathbf{e}^R_{i,s,t}$ through both source-state routing and message transformation.

For each source state $t \in \mathcal{I}_{i,s}$, the routing logit combines latent compatibility between the target and source representations with an explicit relation-dependent term,

$$a^{\mathrm{rel}}_{i,s,t} = \frac{Q(\mathbf{h}_{i,s})^\top K(\mathbf{h}_{i,t})}{\sqrt{d}} + b_\rho(\mathbf{e}^R_{i,s,t}). \tag{21}$$

The first term measures compatibility in the learned state representation, whereas $b_\rho(\mathbf{e}^R_{i,s,t})$ modifies the interaction according to the observed structural relation between the two conformations.

A null source state $\varnothing$ is introduced with a fixed logit $a^{\mathrm{rel}}_{i,s,\varnothing} = 0$ and zero message $\mathbf{m}_{i,s\leftarrow\varnothing} = \mathbf{0}$. Normalization over $\mathcal{I}_{i,s} \cup \{\varnothing\}$ therefore gives

$$\begin{aligned} \alpha^{\mathrm{rel}}_{i,s,t} &= \frac{\exp(a^{\mathrm{rel}}_{i,s,t})}{1 + \sum_{u\in\mathcal{I}_{i,s}} \exp(a^{\mathrm{rel}}_{i,s,u})}, \qquad t \in \mathcal{I}_{i,s}, \\ \alpha^{\mathrm{rel}}_{i,s,\varnothing} &= \frac{1}{1 + \sum_{u\in\mathcal{I}_{i,s}} \exp(a^{\mathrm{rel}}_{i,s,u})}, \\ \sum_{t\in\mathcal{I}_{i,s}} \alpha^{\mathrm{rel}}_{i,s,t} &= 1 - \alpha^{\mathrm{rel}}_{i,s,\varnothing}. \end{aligned} \tag{22}$$

Unlike normalization restricted to the non-null source states, Eq. 22 does not force the imported cross-state information to have unit total mass. The null probability determines how much of the update is assigned to the zero-message alternative, while the remaining probability mass is distributed among the observed source conformations.

The same relation embedding also parameterizes the content transferred from state t to state s. Multiplicative and additive relation terms are defined as $\boldsymbol{\gamma}_{i,s,t} = \tanh(W_\gamma \mathbf{e}^R_{i,s,t})$ and $\boldsymbol{\beta}_{i,s,t} = W_\beta \mathbf{e}^R_{i,s,t}$, yielding

$$\begin{aligned} \mathbf{m}_{i,s\leftarrow t} &= (1 + \boldsymbol{\gamma}_{i,s,t}) \odot W_V \mathbf{h}_{i,t} + \boldsymbol{\beta}_{i,s,t}, \\ \mathbf{c}^{\mathrm{rel}}_{i,s} &= \sum_{t\in\mathcal{I}_{i,s}} \alpha^{\mathrm{rel}}_{i,s,t} \mathbf{m}_{i,s\leftarrow t}, \\ \widetilde{\mathbf{h}}_{i,s} &= \mathbf{h}_{i,s} + \eta_{\mathrm{rel}} \Gamma_\rho\left(\mathbf{c}^{\mathrm{rel}}_{i,s}\right). \end{aligned} \tag{23}$$

Substituting Eqs. 21–23 gives the explicit form of the abstract operator $\mathcal{F}_\rho$ in Eq. 3. The structural relation affects the update through two distinct paths: the bias $b_\rho(\mathbf{e}^R_{i,s,t})$ changes the weight assigned to each source state, while $(\boldsymbol{\gamma}_{i,s,t}, \boldsymbol{\beta}_{i,s,t})$ changes the message contributed by that state.

The null-augmented normalization is especially relevant when only one alternative conformation is available. If $\mathcal{I}_{i,s} = \{t\}$, then

$$\alpha^{\mathrm{rel}}_{i,s,t} = \frac{\exp(a^{\mathrm{rel}}_{i,s,t})}{1 + \exp(a^{\mathrm{rel}}_{i,s,t})} = \sigma\left(a^{\mathrm{rel}}_{i,s,t}\right), \qquad \alpha^{\mathrm{rel}}_{i,s,\varnothing} = 1 - \alpha^{\mathrm{rel}}_{i,s,t}. \tag{24}$$

Thus, for a two-state ensemble, the source-state relation continues to determine the strength of cross-state communication. A softmax over the single non-null source alone would instead assign unit weight independently of its interaction score.

The formulation also contains two direct reductions. If the relation-dependent terms are removed,

$$b_\rho(\mathbf{e}^R_{i,s,t}) = 0, \qquad \boldsymbol{\gamma}_{i,s,t} = \mathbf{0}, \qquad \boldsymbol{\beta}_{i,s,t} = \mathbf{0},$$

Eq. 23 reduces to latent-only cross-state communication,

$$\mathbf{m}_{i,s\leftarrow t} = W_V \mathbf{h}_{i,t}, \qquad a^{\mathrm{rel}}_{i,s,t} = \frac{Q(\mathbf{h}_{i,s})^\top K(\mathbf{h}_{i,t})}{\sqrt{d}}.$$

Setting $\eta_{\mathrm{rel}} = 0$ further yields $\widetilde{\mathbf{h}}_{i,s} = \mathbf{h}_{i,s}$, corresponding to independent state encoding before ensemble aggregation. If $\mathcal{I}_{i,s} = \varnothing$, the null state receives all routing mass and the same identity update follows automatically.

A.4 State-Retentive Decoding and Centered-Residual Properties

RCSC produces relation-updated state representations $\{\widetilde{\mathbf{h}}_{i,s}\}_{s\in\mathcal{S}_i}$. For $n_i = |\mathcal{S}_i|$, SRAD separates their permutation-invariant consensus from their state-indexed deviations:

$$\boldsymbol{\mu}_i = \frac{1}{n_i}\sum_{s\in\mathcal{S}_i}\widetilde{\mathbf{h}}_{i,s}, \qquad \boldsymbol{\delta}_{i,s} = \widetilde{\mathbf{h}}_{i,s} - \boldsymbol{\mu}_i, \qquad \widetilde{\mathbf{h}}_{i,s} = \boldsymbol{\mu}_i + \boldsymbol{\delta}_{i,s}, \qquad \sum_{s\in\mathcal{S}_i}\boldsymbol{\delta}_{i,s} = \mathbf{0}. \tag{25}$$

Hence $\{\widetilde{\mathbf{h}}_{i,s}\} \leftrightarrow (\boldsymbol{\mu}_i, \boldsymbol{\Delta}_i)$ is exactly invertible over the observed scalar states, where $\boldsymbol{\Delta}_i = \{\boldsymbol{\delta}_{i,s}\}_{s\in\mathcal{S}_i}$. The consensus is combined with the pooled vector representation and propagated through the shared geometric pathway. At decoding step t, this produces $\mathbf{q}_{\pi_t} = D_\phi(\mathbf{G}, Y_{\pi_{<t}})_{\pi_t}$, while $\boldsymbol{\Delta}_{\pi_t}$ remains explicitly indexed by state.

The decoder state retrieves residual information through shared projections and the same null-augmented normalization used by RCSC:

$$\begin{aligned} a^{\mathrm{mem}}_{\pi_t,s} &= \frac{(W^m_Q\mathbf{q}_{\pi_t})^\top(W^m_K\boldsymbol{\delta}_{\pi_t,s})}{\sqrt{d_m}}, \\ \beta_{\pi_t,s} &= \frac{\exp(a^{\mathrm{mem}}_{\pi_t,s})}{1+\sum_{u\in\mathcal{S}_{\pi_t}}\exp(a^{\mathrm{mem}}_{\pi_t,u})}, \\ \mathbf{c}^{\mathrm{state}}_{\pi_t} &= W_c\sum_{s\in\mathcal{S}_{\pi_t}}\beta_{\pi_t,s}W^m_V\boldsymbol{\delta}_{\pi_t,s}, \\ \widetilde{\mathbf{q}}_{\pi_t} &= \mathbf{q}_{\pi_t} + \eta_{\mathrm{mem}}\mathbf{c}^{\mathrm{state}}_{\pi_t}. \end{aligned} \tag{26}$$

The null state has a fixed logit of zero and a zero-valued message. The projections W^m_Q, W^m_K, W^m_V, and W_c are shared across states and implemented without additive bias; no state-wise nonlinear transformation is applied to the residual value before aggregation.

The centered representation yields the central structural property of SRAD. If every observed state receives the same retrieval weight, $\beta_{\pi_t,s} = \beta_{\pi_t}$, then

$$\mathbf{c}^{\mathrm{state}}_{\pi_t} = \beta_{\pi_t}W_cW^m_V\sum_{s\in\mathcal{S}_{\pi_t}}\boldsymbol{\delta}_{\pi_t,s} = \mathbf{0}, \qquad \widetilde{\mathbf{q}}_{\pi_t} = \mathbf{q}_{\pi_t}. \tag{27}$$

Thus, uniform state relevance introduces no residual correction: ensemble-shared information is carried by the consensus pathway, whereas the residual pathway responds only to non-uniform state-specific evidence. The zero-correction result relies on the shared bias-free linear value and output maps, which preserve the zero-sum residual structure.

The same result covers the principal boundary cases. If all state representations coincide, every residual is zero; if $n_i = 1$, then $\boldsymbol{\mu}_i = \widetilde{\mathbf{h}}_{i,1}$ and $\boldsymbol{\delta}_{i,1} = \mathbf{0}$. SRAD therefore reduces to the shared autoregressive pathway whenever no state-specific deviation exists.

The state ordering does not affect the retrieved representation. For a permutation P of the conformational axis, the consensus is invariant, $\boldsymbol{\mu}_i(P\widetilde{\mathbf{H}}) = \boldsymbol{\mu}_i(\widetilde{\mathbf{H}})$, whereas the residual collection is permutation equivariant, $\boldsymbol{\Delta}_i(P\widetilde{\mathbf{H}}) = P\boldsymbol{\Delta}_i(\widetilde{\mathbf{H}})$. Shared memory projections induce the same permutation of the scores and retrieval weights, while the weighted summation removes the state ordering. Hence

$$\mathbf{c}^{\mathrm{state}}_{\pi_t}(P\boldsymbol{\Delta}_{\pi_t}) = \mathbf{c}^{\mathrm{state}}_{\pi_t}(\boldsymbol{\Delta}_{\pi_t}), \qquad \widetilde{\mathbf{q}}_{\pi_t}(P\boldsymbol{\Delta}_{\pi_t}) = \widetilde{\mathbf{q}}_{\pi_t}(\boldsymbol{\Delta}_{\pi_t}). \tag{28}$$

Combined with the permutation-equivariant state encoding and RCSC update, the final autoregressive sequence distribution is invariant to the ordering of the conformational states.

Finally, the consensus is invariant and the residual set equivariant under permutations of the conformational axis. Shared projections permute memory scores and weights in the same way, while the state summation removes that ordering. Together with the permutation-equivariant geometric encoder and RCSC update, this gives

$$p_\theta(Y \mid P\mathcal{X}, PM) = p_\theta(Y \mid \mathcal{X}, M)$$

for any permutation P of the state axis.

B Additional Experimental Details

B.1 Dataset Details

We adopt the multi-conformational protein benchmark introduced by DynamicMPNN (Abrudan et al., 2026). The benchmark is derived from experimentally resolved structures in CoDNaS (`v2025`) (Monzon et al., 2016) and the Protein Data Bank (PDB). Protein chains are clustered at $\geq 95\%$ sequence identity, yielding 46,033 clusters with at least two observed conformations. For each cluster, the pair with the largest inter-state RMSD is selected to represent structurally distinct conformations. AlphaFold-predicted structures are not included.

We use the original benchmark splits, comprising 44,243 multi-conformational training pairs, 100 validation proteins, and 96 test proteins. The training set is filtered against the evaluation proteins according to the benchmark's sequence- and structure-similarity criteria. An additional 27,394 single-state sequence–structure pairs from non-overlapping clusters are included in training. The experiments use the dataset splits described above.

B.2 Evaluation Metrics Details

B.2.1 Sequence-Level Metrics

Sequence recovery. Following standard inverse-folding evaluation (Dauparas et al., 2022; Abrudan et al., 2026), we report sequence recovery as the fraction of correctly predicted residues relative to the native sequence,

$$\text{Recovery} = \frac{1}{L}\sum_{i=1}^{L}\mathbb{I}\left[\hat{y}_i = y_i\right], \tag{29}$$

where L is the sequence length, y_i is the native residue, and $\hat{y}_i$ denotes the predicted residue at position i.

Perplexity. For autoregressive models, we additionally report perplexity computed from the native sequence likelihood,

$$\text{PPL} = \exp\left(-\frac{1}{L}\sum_{i=1}^{L}\log p_\theta(y_i \mid y_{<i}, \mathcal{X})\right). \tag{30}$$

Lower perplexity indicates a higher probability assigned to the native sequence under the target conformational ensemble.

B.2.2 Structural Self-Consistency

Sequence recovery does not directly measure whether a designed sequence remains compatible with all target conformations. Following DynamicMPNN (Abrudan et al., 2026), we therefore evaluate structural self-consistency using AlphaFold3 (AF3) (Abramson et al., 2024). For a designed sequence Y and target state $X^{(s)}$, AF3 is run with $X^{(s)}$ supplied as a structural template, producing a predicted structure $\widehat{X}^{(s)}$.

Template-conditioned structure prediction is used as a constrained sequence–structure compatibility test rather than as an unconstrained conformational search procedure. Previous work has investigated AlphaFold-derived scores for assessing candidate protein structures (Roney & Ovchinnikov, 2022). We report the following structural metrics.

pLDDT. We use the mean AF3-predicted local distance difference test (pLDDT) as a confidence measure for each predicted structure. pLDDT is a per-residue prediction of local structural accuracy derived from the lDDT framework (Jumper et al., 2021; Mariani et al., 2013); higher values indicate greater confidence in the predicted local geometry.

RMSD. After structurally aligning the predicted and target conformations, we compute the Cα root-mean-square deviation,

$$\text{RMSD} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\|\widehat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2}, \tag{31}$$

Table 6: **Template-conditioned versus template-free AF3 evaluation.** The same sequence is evaluated against every target conformation. For template-free AF3, one shared ensemble of K predictions per sequence is matched to all targets; for template-conditioned AF3, each target is supplied as its own template. Worst-state TM is calculated within each designed sequence before aggregation. SR@0.8 is the fraction of designed sequences for which every target state is recovered with TM-score ≥ 0.8.

	Target-template AF3			Template-free AF3		
Method	**Mean TM ↑**	**Worst TM ↑**	**SR@0.8 (%) ↑**	**Mean TM ↑**	**Worst TM ↑**	**SR@0.8 (%) ↑**
DynamicMPNN	0.872	0.781	46.3	0.797	0.692	12.4
ReStateMPNN	**0.907**	**0.832**	**51.6**	**0.825**	**0.743**	**18.9**

where M is the number of aligned residues. Lower RMSD indicates closer agreement between the predicted and target structures.

TM-score. We use TM-score (Zhang & Skolnick, 2004) to measure global structural similarity after optimal alignment. Unlike RMSD, TM-score is normalized for protein length and is less sensitive to isolated large local deviations. Structural alignments are obtained using TM-align (Zhang & Skolnick, 2005). Higher TM-score indicates greater global structural agreement.

lDDT. For residue-level structural agreement, we additionally report lDDT where applicable. lDDT evaluates local interatomic distance preservation without requiring global superposition and is therefore complementary to global metrics such as RMSD and TM-score (Mariani et al., 2013).

B.2.3 Decoy-Normalized AF3 Self-Consistency

Template-conditioned AF3 predictions may partially follow the supplied template even when the designed sequence is not specifically compatible with that structure. To compare target-template performance with a structurally dissimilar template, we use the decoy-normalization procedure described by DynamicMPNN (Abrudan et al., 2026). For each target state $X^{(s)}$, we select a structurally dissimilar decoy D satisfying $\mathrm{TM}(X^{(s)}, D) < 0.4$ and evaluate the same designed sequence under both target and decoy templates.

For a metric m for which larger values are better, such as pLDDT or TM-score, we define

$$m_{\mathrm{norm}} = \frac{m\big(\mathrm{AF3}(Y, X^{(s)}), X^{(s)}\big)}{m(\mathrm{AF3}(Y, D), D)}. \tag{32}$$

Values above one indicate that the corresponding metric is higher under the target template than under the selected decoy template.

For RMSD, where smaller values are preferable, we use

$$\mathrm{RMSD}_{\mathrm{norm}} = \frac{\mathrm{RMSD}\left(\mathrm{AF3}(Y, X^{(s)}), X^{(s)}\right)}{\mathrm{RMSD}\left(\mathrm{AF3}(Y, D), D\right)}. \tag{33}$$

Lower values therefore indicate that the designed sequence refolds more closely to the intended target state than to an unrelated structural template.

The normalized metrics should be interpreted as relative sequence–structure specificity scores rather than direct estimates of physical free-energy differences. Their purpose is to distinguish target-specific structural compatibility from generic template-following behavior.

B.2.4 Multi-State Aggregation

For a designed sequence evaluated against $\mathcal{X} = \{X^{(1)}, \ldots, X^{(N)}\}$, each structural metric is computed independently for every target state. We report both ensemble-average and worst-state performance.

For metrics for which larger values are better, such as TM-score and pLDDT,

$$m_{\mathrm{mean}} = \frac{1}{N}\sum_{s=1}^{N} m_s, \qquad m_{\mathrm{worst}} = \min_s m_s. \tag{34}$$

Table 7: Key implementation and training configurations of ReStateMPNN.

Configuration	Value
Trainable parameters	4.55M
Node hidden dims. (scalar/vector)	128 / 16
Edge hidden dims. (scalar/vector)	32 / 4
State encoder / pooled encoder / decoder	8 / 4 / 4 GVP layers
KNN neighbors	32
Inter-state attention heads	4
Relation embedding dimension	16
State-memory heads	4
Dropout	0.5
Optimizer	Adam
Initial learning rate	1×10^{-3}
Weight decay / label smoothing	0 / 0.05
Effective batch size	32
Precision	BF16 mixed
Early stopping patience	10
Random seed	52

For RMSD, where lower values indicate better structural agreement,

$$\mathrm{RMSD}_{\mathrm{mean}} = \frac{1}{N}\sum_{s=1}^{N}\mathrm{RMSD}_s, \qquad \mathrm{RMSD}_{\mathrm{worst}} = \max_s \mathrm{RMSD}_s. \tag{35}$$

Worst-state metrics are particularly relevant to multi-state inverse folding because they penalize designs that are compatible with one target conformation but fail to recover another. Unless otherwise specified, reported benchmark results are averaged over all sampled sequences and conformational states using the same evaluation protocol across models.

B.3 Implementation Details

Architecture and feature representation. ReStateMPNN builds on the GVP-based multi-state inverse-folding architecture of DynamicMPNN (Abrudan et al., 2026). Protein structures are represented as residue-level Cα graphs with k-nearest-neighbor edges ($k = 32$). We use 27/2 scalar/vector node input channels and 18/1 scalar/vector edge input channels, with hidden dimensions of 128/16 and 32/4, respectively. Each conformation is encoded by an eight-layer GVP encoder, followed by a four-layer pooled encoder and a four-layer autoregressive GVP decoder.

The proposed inter-state block uses four attention heads and a 16-dimensional relation embedding constructed from pairwise TM similarity, structural overlap, contact Jaccard similarity, and contact gain/loss features. State-specific residuals are retained through a four-head memory module with gated residual fusion. The complete model contains 4.55M trainable parameters. Unless otherwise specified, ablation variants share the same backbone and differ only in the corresponding relation or state-memory component. We follow DynamicMPNN for geometric node and edge features and apply Gaussian coordinate noise with standard deviation 0.1 Å during training.

Training. We use a two-stage combined pretraining and Multi-Finetuning procedure. Combined pretraining samples single-state and multi-state examples with equal probability, while multi-state finetuning is initialized from the best pretraining checkpoint and uses only multi-state examples. Both stages optimize autoregressivecross-entropy with Adam, an initial learning rate of 10^{-3}, no weight decay, and label smoothing of 0.05. Training uses BF16 mixed precision and distributed data parallelism. Models are evaluated after each epoch, and the checkpoint with the highest validation sequence recovery is retained. All reported experiments use random seed 52.

C Limitations and Future Work

ReStateMPNN is formulated around a finite set of aligned conformational states and focuses on sequence compatibility with prescribed structural alternatives. While this setting provides a practical

abstraction for multi-state inverse folding, it does not capture the full complexity of conformational ensembles or their underlying energetic and dynamical organization. Future work may extend the current geometry-based relation modeling toward richer ensemble representations when corresponding structural or biophysical information is available.

The present study evaluates designed sequences computationally through sequence recovery and structure-based refolding metrics. These measurements characterize consistency with the target conformations but do not replace experimental validation of multi-state behavior. We therefore view experiment-in-the-loop design as an important extension, where measurements from designed proteins are iteratively incorporated as feedback for subsequent design rounds, providing a path toward tighter alignment between computational objectives and experimentally observed conformational behavior.

D LLM USAGE STATEMENT

The authors acknowledge the use of LLMs in the preparation of this manuscript. LLMs were employed as writing assistants to refine language, improve grammar, and provide stylistic suggestions, and to help verify the notational consistency of mathematical formulas.

E REPRODUCIBILITY STATEMENT

To facilitate reproducibility and independent verification, we provide an anonymized, publicly accessible code repository at https://anonymous.4open.science/r/ReStateMPNN/. The repository includes the complete ReStateMPNN implementation, environment and dependency specifications, execution scripts, testing modules, and model weights.

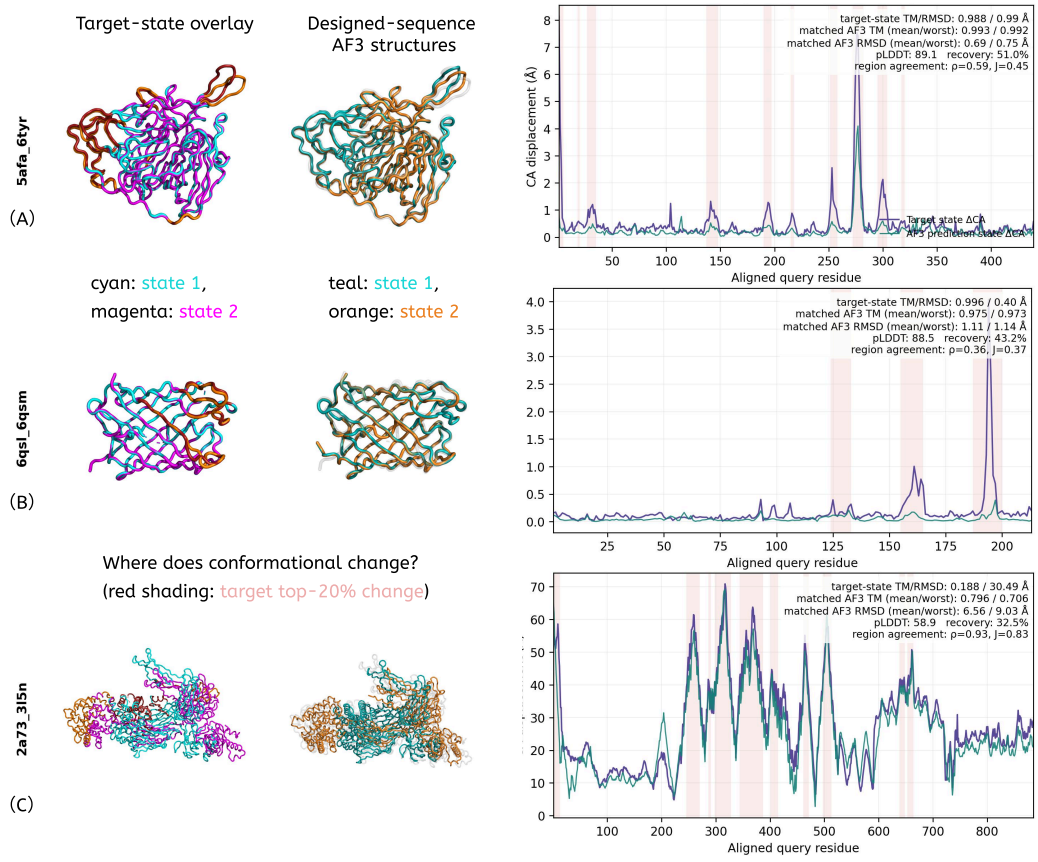

Figure 3: **Additional structural case analysis.** Target-state structures are shown in cyan and magenta, while AF3 structures obtained from the designed sequence are shown in teal and orange. The right panels compare residue-wise Cα displacement between the two target states and the corresponding AF3 structures. Red shading denotes residues in the top 20% of target-state displacement. Reported statistics include matched-state TM-score and RMSD, pLDDT, sequence recovery, displacement-profile correlation ρ, and Jaccard overlap J of high-displacement regions.

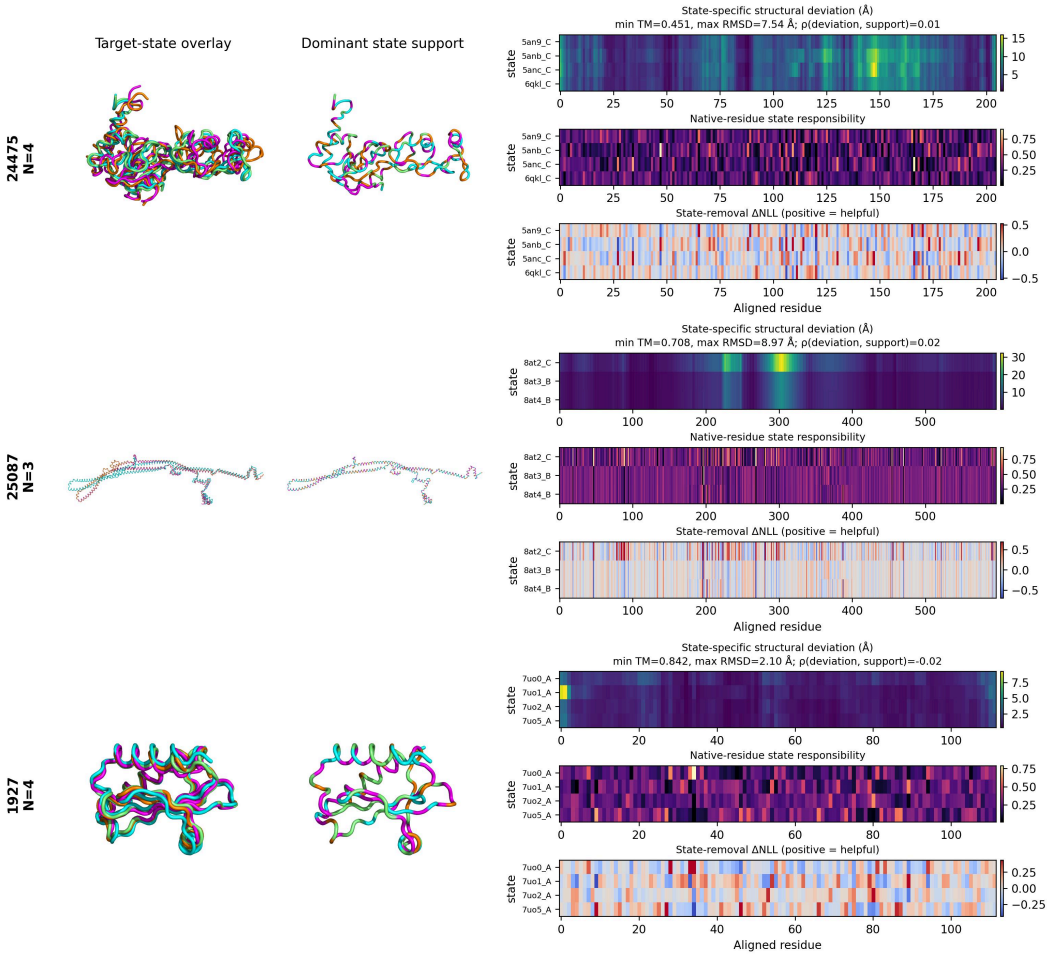

Figure 4: **State-support diagnostics for multi-state ensembles with $N > 2$.** Each row shows the target-state ensemble, the dominant state support, state-specific structural deviation, native-residue state responsibility, and state-removal ΔNLL. State responsibility is a normalized support diagnostic rather than a learned attention weight. State-removal ΔNLL is computed under teacher forcing; positive values indicate that removing a state decreases the likelihood of the native residue.

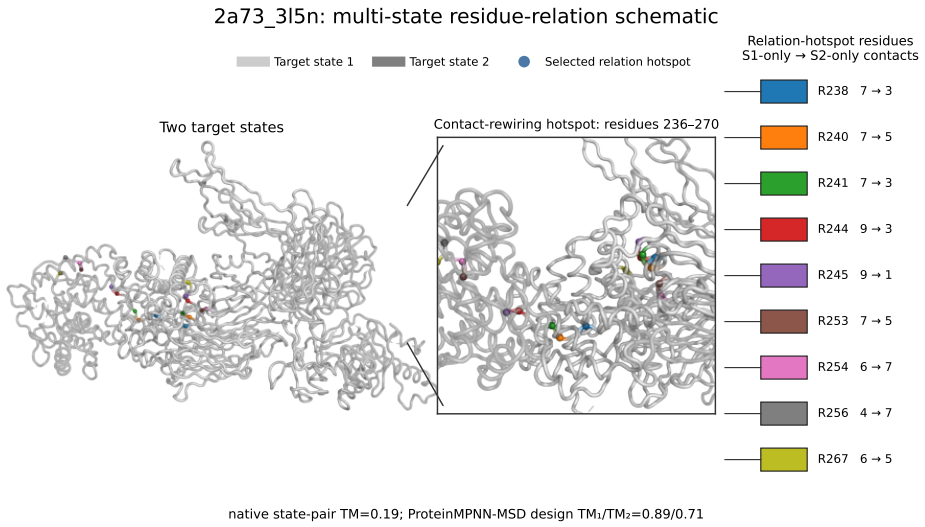

Figure 5: **Residue-relation schematic for `2a73_3l5n`.** Target state 1 and state 2 are shown in light and dark gray, respectively. Blue dots mark selected relation hotspots. The inset enlarges the contact-rewiring hotspot (residues 236–270). The list on the right reports the number of contacts for each hotspot residue in state 1 → state 2.

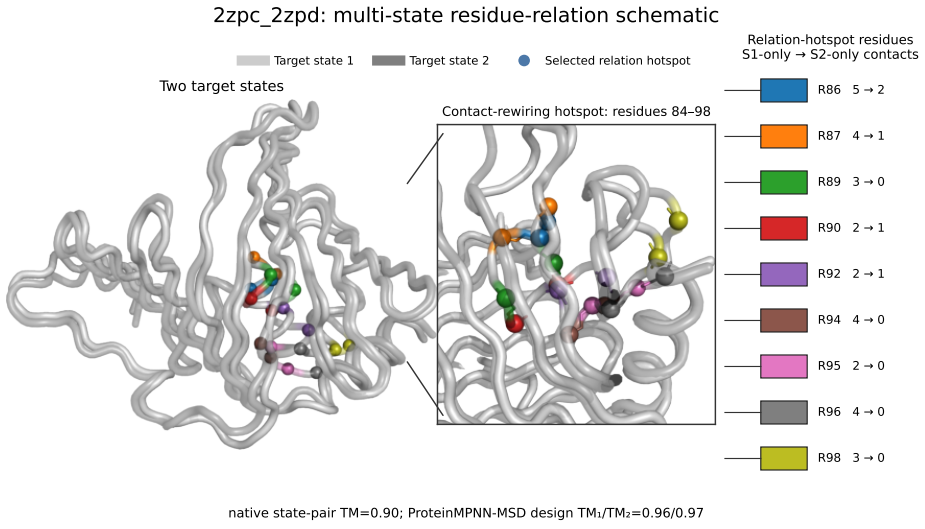

Figure 6: **Residue-relation schematic for `2zpc_2zpd`.** Target state 1 and state 2 are shown in light and dark gray, respectively. Blue dots mark selected relation hotspots. The inset enlarges the contact-rewiring hotspot (residues 84–98). The list on the right reports the number of contacts for each hotspot residue in state 1 → state 2. Native state-pair TM = 0.90; ProteinMPNN-MSD design TM_1/TM_2 = 0.96 / 0.97.

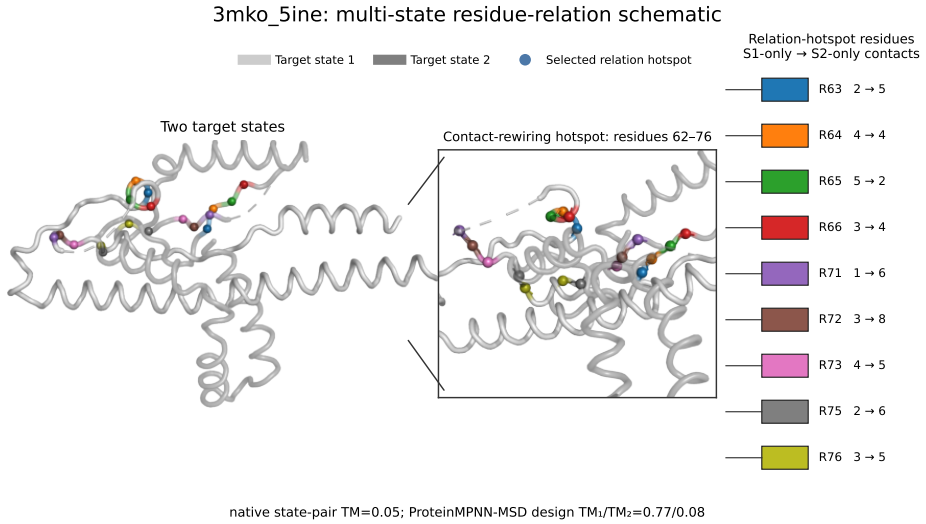

Figure 7: **Residue-relation schematic for `3mko_5ine`.** Target state 1 and state 2 are shown in light and dark gray, respectively. Blue dots mark selected relation hotspots. The inset enlarges the contact-rewiring hotspot (residues 62–76). The list on the right reports the number of contacts for each hotspot residue in state 1 → state 2.

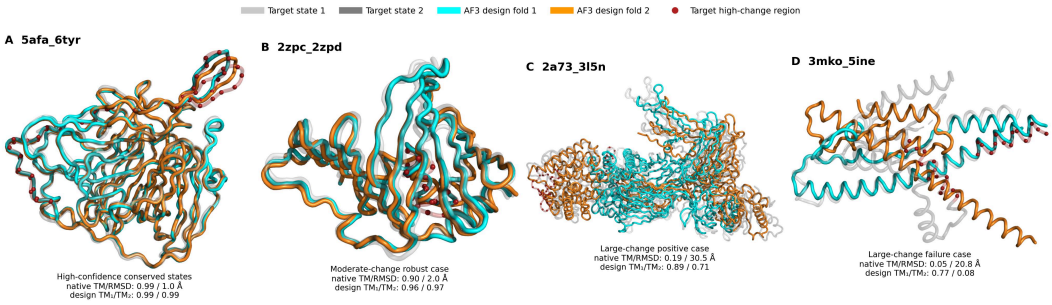

Figure 8: **AlphaFold3 state retrieval on designed sequences across conformational change scales.** Experimental target states are shown in light and dark gray (state 1 and 2), while AlphaFold3 structures of the same designed sequence conditioned on the target states are shown in cyan (design fold 1) and orange (design fold 2). Red markers identify the contiguous target region with the largest mean inter-state Cα displacement. The cases span diverse regimes: (A) `5afa_6tyr` (conserved states, native TM/RMSD: 0.99 / 1.0 Å, design TM_1/TM_2: 0.99 / 0.99), (B) `2zpc_2zpd` (moderate-change robust case, native TM/RMSD: 0.90 / 2.0 Å, design TM_1/TM_2: 0.96 / 0.97), (C) `2a73_3l5n` (large-change positive case, native TM/RMSD: 0.19 / 30.5 Å, design TM_1/TM_2: 0.89 / 0.71), and (D) `3mko_5ine` (large-change failure case, native TM/RMSD: 0.05 / 20.8 Å, design TM_1/TM_2: 0.77 / 0.08).

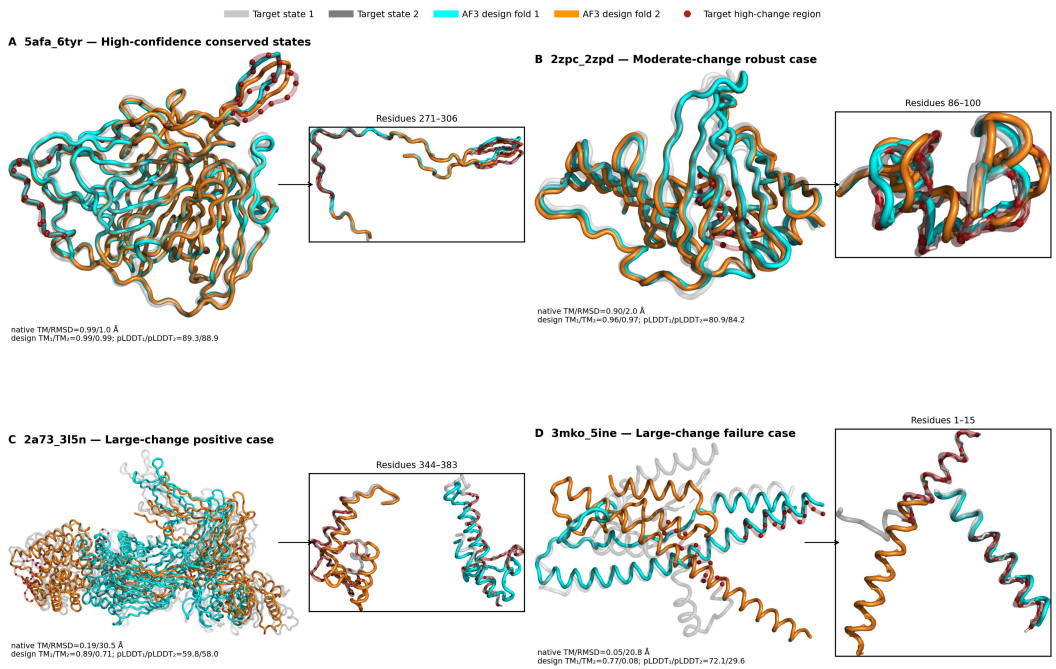

Figure 9: **Whole-structure and regional comparison of ProteinMPNN$_{\mathrm{MSD}}$ designs.** Each panel shows the two experimental states (Target state 1 in light gray; Target state 2 in dark gray) and the corresponding AlphaFold3 predictions (cyan, design fold 1; orange, design fold 2) generated from a single ProteinMPNN-MSD-designed sequence. Red markers denote the target high-change region. Insets enlarge the contiguous region with the highest mean target-state Cα displacement. (A) `5afa_6tyr` exhibits nearly identical target states and near-native recovery in both conditions (native TM/RMSD: 0.99 / 1.0 Å; design TM$_1$/TM$_2$: 0.99 / 0.99). (B) `2zpc_2zpd` retains high compatibility under moderate local rearrangement (native TM/RMSD: 0.90 / 2.0 Å; design TM$_1$/TM$_2$: 0.96 / 0.97). (C) Despite a native state-pair TM-score of 0.19 and RMSD of 30.5 Å, `2a73_3l5n` recovers both target-conditioned conformations with TM-scores of 0.89 and 0.71. (D) `3mko_5ine` has a state-2 target-template AF3 prediction with a TM-score of 0.08 (native TM/RMSD: 0.05 / 20.8 Å). These structures were obtained with target-template AlphaFold3 evaluation and therefore demonstrate template-conditioned compatibility rather than template-free bistability.

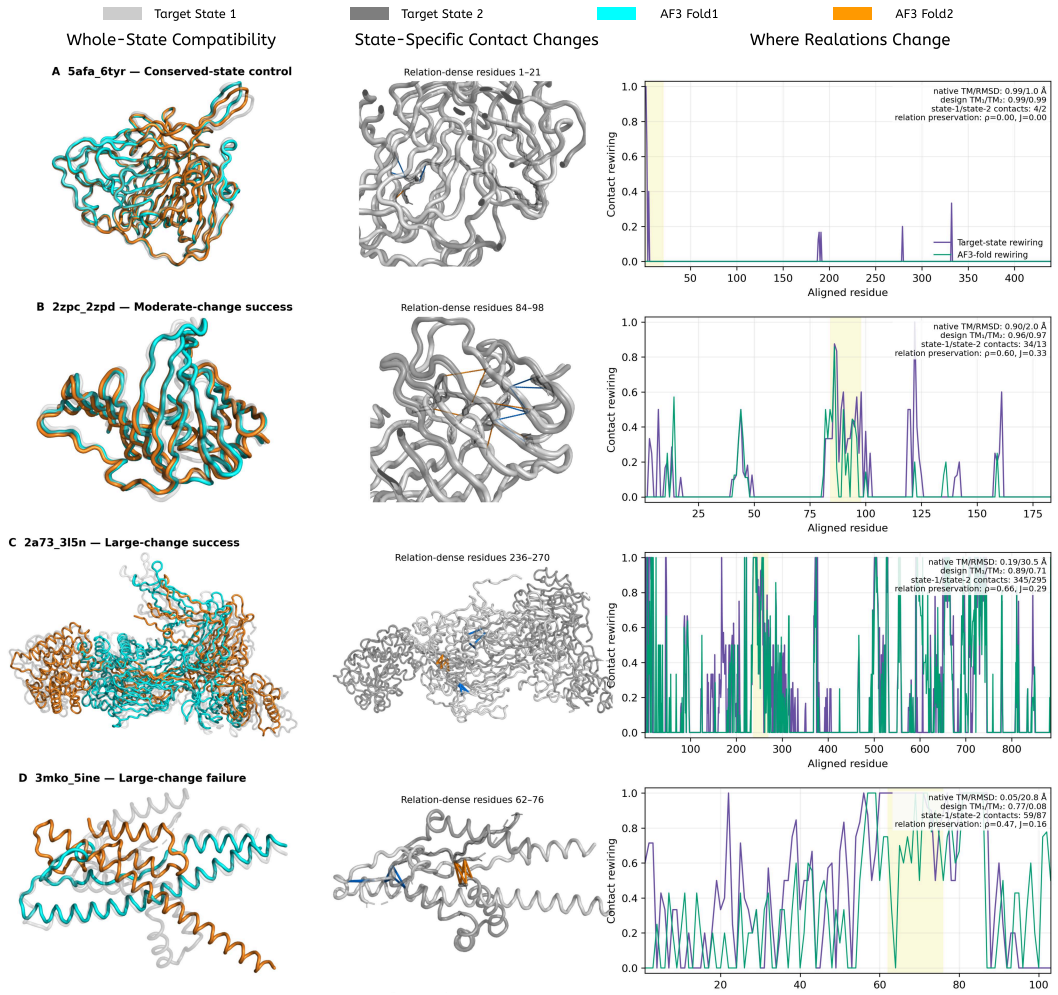

Figure 10: **Whole-structure and residue-relation comparison across multi-state design cases.** Each row shows (left) whole-state compatibility between the two experimental target states (Target state 1 in light gray; Target state 2 in dark gray) and AlphaFold3 predictions from a single designed sequence (cyan, AF3 fold 1; orange, AF3 fold 2), (middle) state-specific contact changes for the relation-dense region, and (right) contact-rewiring profiles along the aligned sequence. Purple lines denote target-state rewiring; green lines denote AF3-fold rewiring. The shaded band marks the relation-dense hotspot. (A) `5afa_6tyr` (conserved-state control; native TM/RMSD: 0.99 / 1.0 Å; design TM_1/TM_2: 0.99 / 0.99; relation preservation: $p = 0.00$, $J = 0.00$). (B) `2zpc_2zpd` (moderate-change success; native TM/RMSD: 0.90 / 2.0 Å; design TM_1/TM_2: 0.96 / 0.97; relation preservation: $p = 0.60$, $J = 0.33$). (C) `2a73_3l5n` (large-change success; native TM/RMSD: 0.19 / 30.5 Å; design TM_1/TM_2: 0.89 / 0.71; relation preservation: $p = 0.66$, $J = 0.29$). (D) `3mko_5ine` (large-change failure; native TM/RMSD: 0.05 / 20.8 Å; design TM_1/TM_2: 0.77 / 0.08; relation preservation: $p = 0.47$, $J = 0.16$).