A Technical Appendices and Supplementary Material

A.1 Equations for Conformational Energy Landscape Overlap Analysis

To quantify the similarity between the protein conformations generated by AI-based models and those in the ProteinConformers dataset, the following three commonly used overlap metrics are employed: Interaction overlap, coverage, and the Jaccard index. These metrics evaluate the extent of agreement in low-energy regions between the protein conformers from different models of the same protein, based on a specified energy threshold.

Let $A = \{A_{i,j}\}$ and $B = \{B_{i,j}\}$ where $i, j \in [0, N]$, denote the two-dimensional free energy landscapes corresponding of two conformational ensembles. Each element $A_{i,j}$ and $B_{i,j}$ represents the free energy value at a specific grid point in the conformational energy landscape. For a given energy threshold τ (e.g., 40 kJ/mol), the number of shared low-energy conformations is defined as:

$$|A \cap B| = \sum_{i,j=1}^{N} \mathbf{1}[A_{i,j} < \tau \wedge B_{i,j} < \tau]$$

where $N = 63$, and $\mathbf{1}[\cdot]$ is the indicator function, which returns 1 if the condition inside is true and 0 otherwise.

The low energy area of different conformational free energy landscape under different threshold are given by:

$$|A| = \sum_{i,j=1}^{N} \mathbf{1}[A_{i,j} < \tau], \qquad |B| = \sum_{i,j=1}^{N} \mathbf{1}[B_{i,j} < \tau]$$

Using the above definitions, the overlap metrics are computed as follows:

- **Interaction**:

$$\text{Interaction} = |A \cap B|$$

- **Coverage** (proportion of low-energy conformations in A also found in B):

$$\text{Coverage} = \frac{|A \cap B|}{|A|}$$

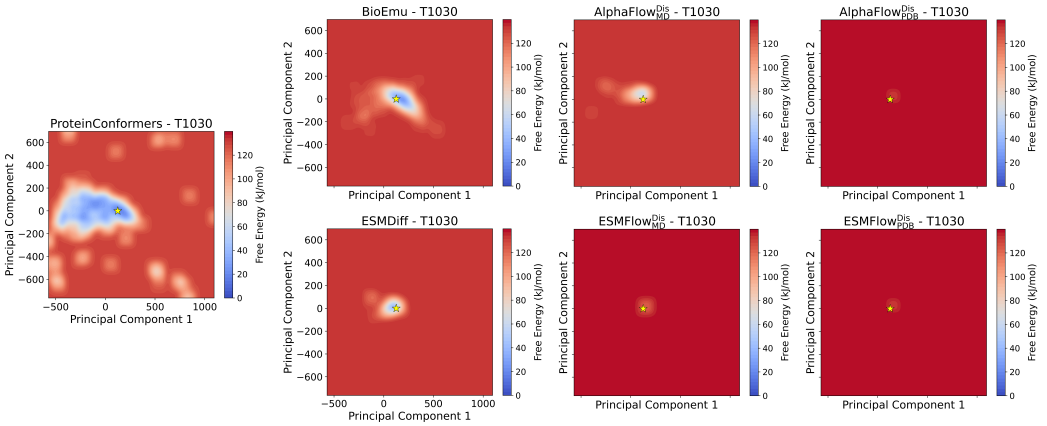
- **Jaccard Index** (symmetric overlap metric between both sets):

$$\text{Jaccard} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

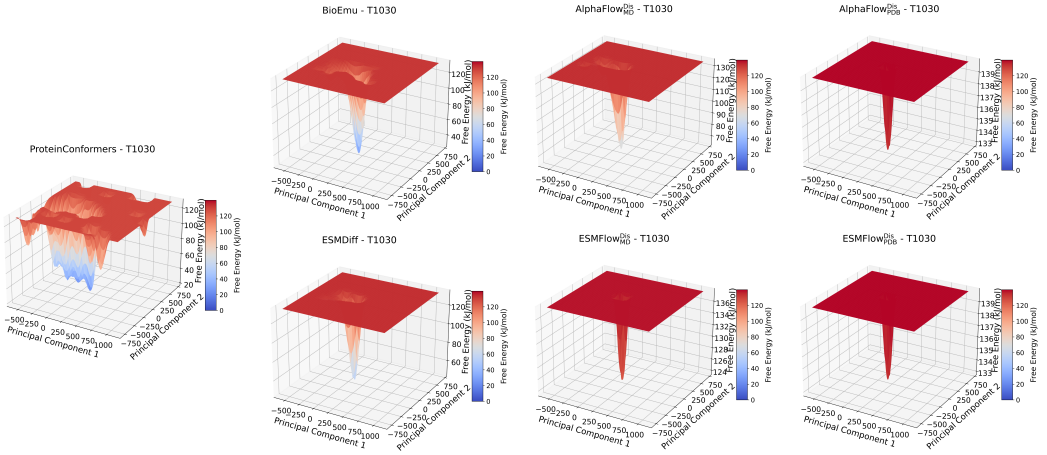
A.2 Free Energy Landscapes Comparison

This section provides additional figures comparing the conformational landscapes of protein conformers from ProteinConformers with those generated by AI models.

A.3 Overview the ProteinConformers proteins

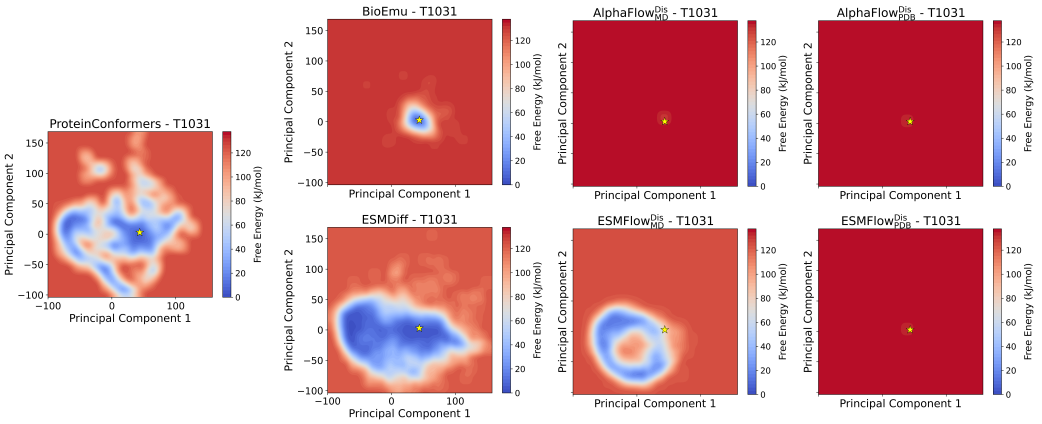


(a) Comparison of 2D conformational landscapes.

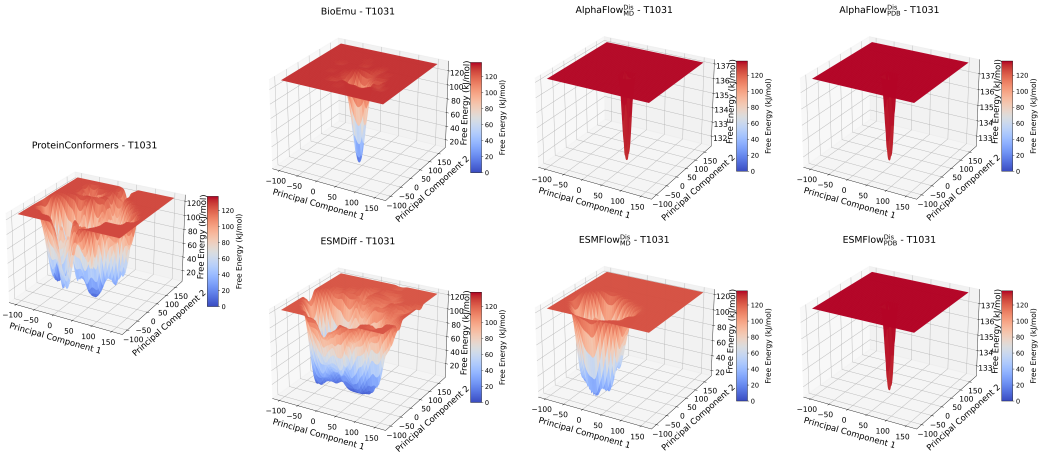


(b) Comparison of 3D conformational landscapes.

Figure 6: Comparison of conformational landscapes for protein T1030, generated by ProteinConformers and protein conformation generative models.

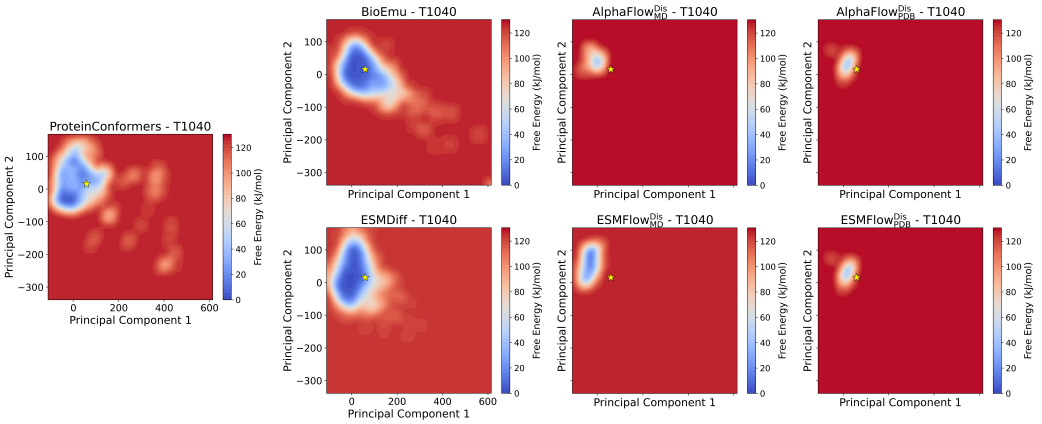


(a) Comparison of 2D conformational landscapes.

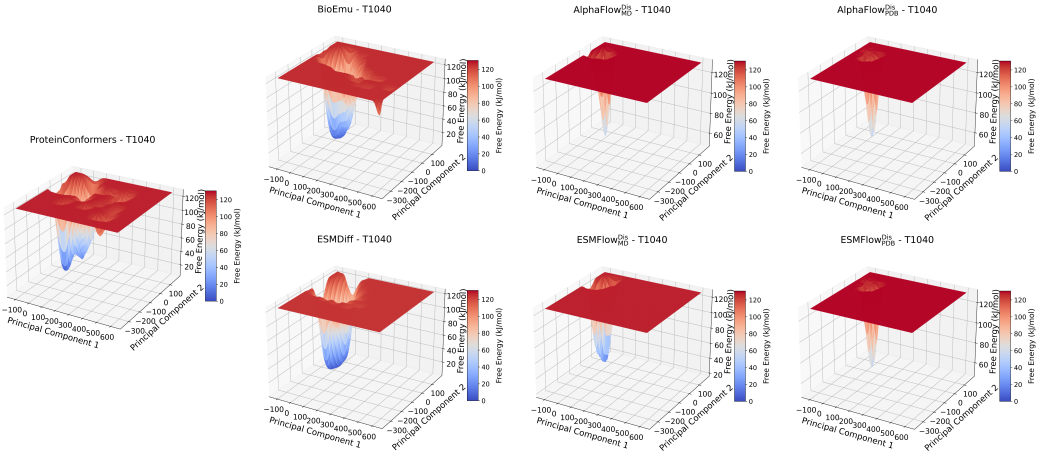


(b) Comparison of 3D conformational landscapes.

Figure 7: Comparison of conformational landscapes for protein T1031, generated by ProteinConformers and protein conformation generative models.

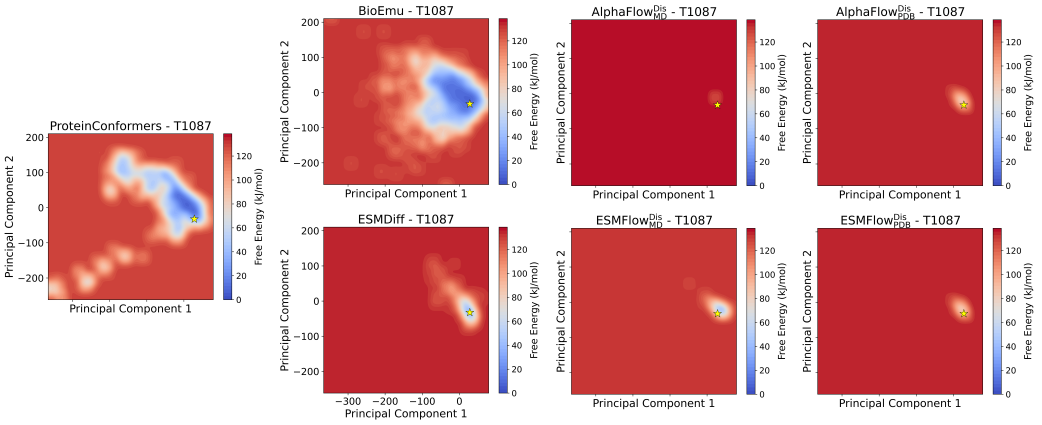


(a) Comparison of 2D conformational landscapes.

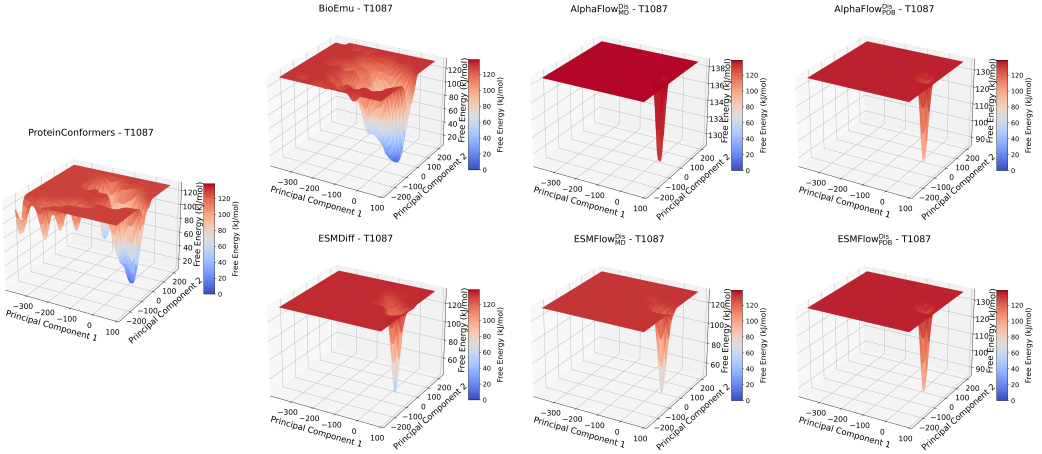


(b) Comparison of 3D conformational landscapes.

Figure 8: Comparison of conformational landscapes for protein T1040, generated by ProteinConformers and protein conformation generative models.



(a) Comparison of 2D conformational landscapes.



(b) Comparison of 3D conformational landscapes.

Figure 9: Comparison of conformational landscapes for protein T1087, generated by ProteinConformers and protein conformation generative models.

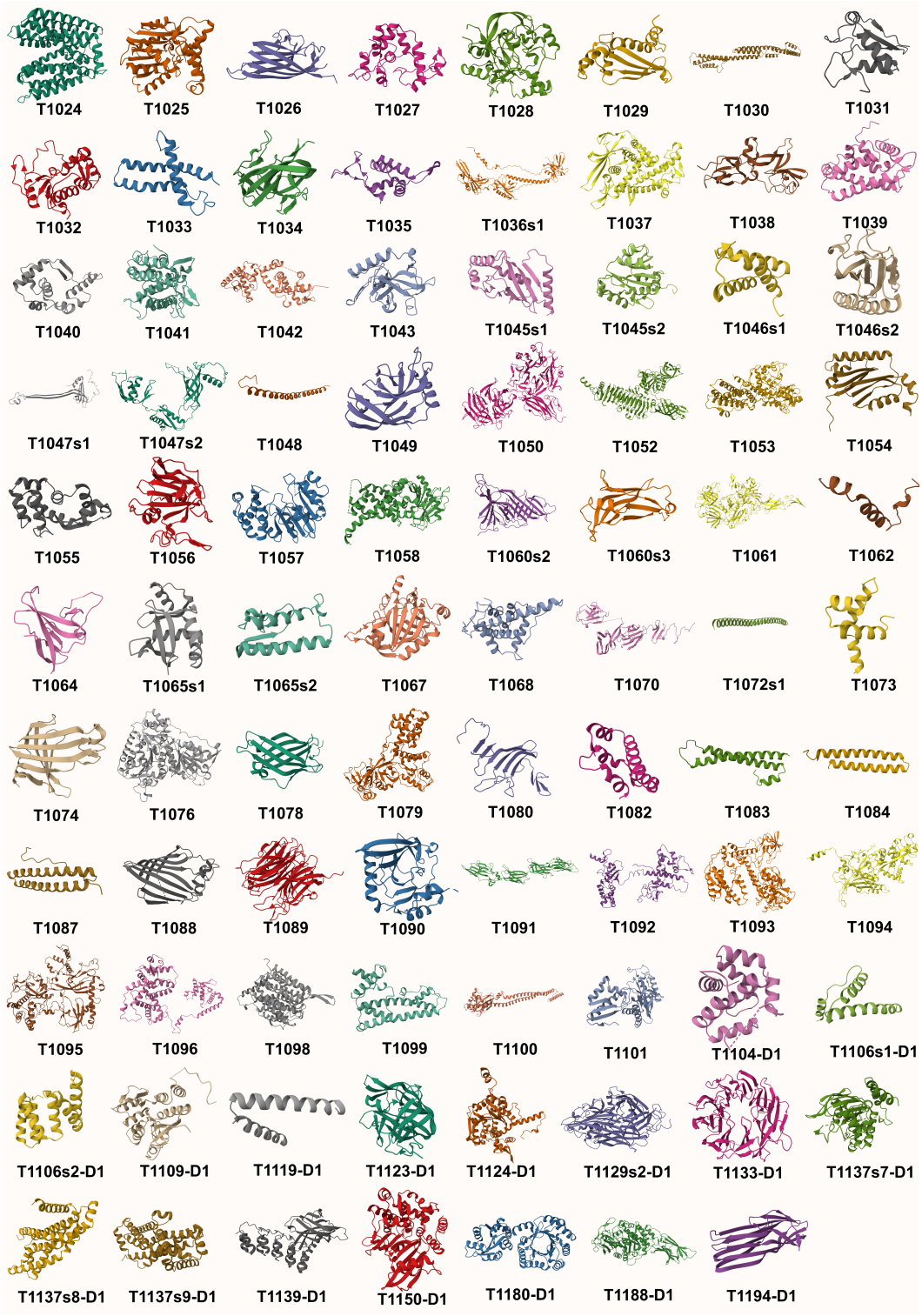


Figure 10: The 3D native structures of all 87 proteins in ProteinConformers.