

Data and text mining

SurfFold: a unified model for protein inverse folding by integrating surface and structural information

Darong Li^{1,†}, Lian Shen^{1,†}, Meijia Song¹, Deyi Li¹, Juan Liu², Xiangrong Liu^{1,3,*} 

¹Department of Computer Science and Technology, Xiamen University, Xiamen, 361005, China

²Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University, Xiamen, 361005, China

³State Key Laboratory of Vaccines for Infectious Diseases, Xiang An Biomedicine Laboratory, Xiamen, 361005, China

*Corresponding author. Department of Computer Science and Technology, Xiamen University, No. 422 Siming South Road, Siming District, Xiamen, Fujian, 361005, China. E-mail: xrlu@xmu.edu.cn.

[†] = Darong Li and Lian Shen equal contribution.

Associate Editor: Jonathan Wren

Abstract

Motivation: Proteins play a crucial role in biological systems, and accurate protein sequence prediction is essential for applications such as drug discovery. Existing inverse folding models primarily rely on protein backbone structure information, overlooking the biochemical properties embedded in protein surface data that constrain its functionality, leading to limited prediction accuracy.

Results: We propose a novel inverse folding framework, SurfFold, which integrates both protein backbone structure and surface information for sequence prediction. Additionally, it incorporates side-chain structural information and its interaction with surface information. Then, we introduce a Representation alignment module to better fuse structure and surface Representations. Experimental results demonstrate that SurfFold achieves state-of-the-art performance on the CATH4.2 dataset, and additional experiments validate the effectiveness of the proposed modules. Moreover, the homologous structure inverse folding experiment also demonstrates that SurfFold possesses excellent capability in homologous protein design.

Availability and implementation: The source code and data are available at <https://github.com/jjudizhengf/SurfFold>.

1 Introduction

Proteins play an indispensable role in biological systems, including providing structural support, catalyzing metabolic reactions, mediating molecular transport, transmitting cellular signals, and supporting immune defense. Therefore, precise protein design is crucial for advancing applications in drug discovery, biocatalysis, materials science, agricultural biotechnology, and environmental remediation (Huang *et al.* 2016). Protein design generally comprises two core aspects: structure design and sequence design. In recent years, considerable progress has been achieved in predicting protein structures from amino acid sequences (Senior *et al.* 2020, Jumper *et al.* 2021). Models such as AlphaFold (Jumper *et al.* 2021) and ESMFold (Lin *et al.* 2023) have significantly improved the accuracy and efficiency of protein tertiary structure prediction. Despite these advances, the inverse problem of predicting amino acid sequences from known protein structures, known as inverse folding, remains a persistent challenge.

Traditional inverse folding methods often rely on energy functions to directly model the protein folding state. While these approaches reflect the underlying physical principles, they are computationally expensive and heavily dependent on expert prior knowledge (Onuchic *et al.* 1997, Dill and MacCallum 2012). In recent years, deep learning techniques have been widely applied to inverse folding tasks, significantly

improving predictive performance (Gao *et al.* 2023b, 2024b, Yang *et al.* 2023). For example, Ingraham *et al.* (2019) developed the GraphTrans model, which enables bidirectional prediction of protein structures and sequences using a graph attention network. Gao *et al.* (2022) proposed AlphaDesign, which integrates graph neural networks with the AlphaFold (Jumper *et al.* 2021) structure prediction tool to generate sequences given a target structure. Zheng *et al.* (2023) introduced LM-design, incorporating structural adapters into protein language models to effectively integrate 3D structure information. Gao *et al.* (2023a) also designed PiFold, which combines innovative feature extractors with a graph neural network architecture to enable one-shot generation of protein sequences. In addition, models such as ProteinMPNN (Dauparas *et al.* 2022), GVP (Jing *et al.* 2020), GCA (Tan *et al.* 2022), ESM-IF (Hsu *et al.* 2022), and ScFold (Zhong *et al.* 2025) have shown strong performance in inverse folding tasks. However, these methods focus primarily on protein structure information while neglecting biochemical information. In homologous protein design, using only structure templates overlooks how protein surface biochemistry affects function. This leads to suboptimal results (Bertoni *et al.* 2017). Crucially, biochemical information on protein surfaces plays a vital role in protein functionality.

In the other field of protein representation learning, researchers have successfully enhanced prediction tasks by

employing multi-modal protein representations that integrate multi-modal information of proteins. For instance, [Yuan et al. \(2024\)](#) introduced the ProteinF3S model, which combines sequence, structure, and surface information to improve the accuracy of enzyme function prediction. [Xu et al. \(2024\)](#) developed the MFE framework, utilizing multi-modal feature extraction to enhance the precision of protein-ligand affinity prediction. Additional methods, including DeepProSite ([Fang et al. 2023](#)), EquiPNAS ([Roche et al. 2024](#)), and GeoBind ([Li and Liu 2023](#)), further demonstrate the effectiveness of integrating multiple modalities for protein feature modeling.

This multi-modal paradigm is particularly critical for the inverse folding task, where the objective is to design protein sequences that fold into specific 3D structures and fulfill intended functional roles. Achieving functional accuracy necessitates capturing not only the geometric constraints imposed by the backbone structure but also integrating the biochemical constraints encoded on the protein surface, which govern molecular interactions and functionality ([Song et al. 2024](#)). Although protein surface representations inherently contain rich biochemical information ([Song et al. 2024](#)) and are derived from the underlying 3D structure, naively combining discrete structure representations with continuous surface representations often results in spatial reference frame misalignment and semantic inconsistencies between geometric and biochemical representations. This spatial and semantic misalignment hinders the model’s ability to learn a unified representation that faithfully reflects the intricate coupling between structure and surface biochemistry. The SurfPro model ([Song et al. 2024](#)), e.g., demonstrated the value of surface biochemical attributes for sequence prediction but did not explicitly integrate geometric structure representations, leaving the challenge of reconciling these two critical modalities unresolved.

To effectively bridge this gap and construct a cohesive multi-modal representation for inverse folding, we leverage side-chain conformations. Side-chains reside precisely at the interface between structure and surface: their geometric conformations are constrained by the backbone scaffold, while their chemical identities and spatial arrangements directly define the physicochemical properties of the protein surface ([Zhu et al. 2012](#)). Moreover, these conformations are profoundly influenced and shaped by the local biochemical environment ([Rose 2019](#)). Consequently, utilizing side-chain information provides a natural and biophysically grounded mechanism to integrate geometric structural constraints with surface biochemical information, thereby establishing a more expressive and functionally predictive framework for protein sequence design.

Motivated by these considerations, we propose SurfFold, a novel inverse folding framework that integrates protein structural and surface representations. We first adapt the three-level feature extraction strategy introduced in ProNet ([Wang et al. 2022](#)) to obtain backbone and side-chain structural information. We then generate the protein surface using the dMaSIF ([Sverrisson et al. 2021](#)) framework. Our architecture includes three main components: a feature encoding module, a representation alignment module, and a one-shot decoding module. The backbone and side-chain information are jointly encoded using E(n)-equivariant graph neural networks (EGNNs) ([Satorras et al. 2021](#)), while surface features are processed using geodesic convolution. Spatial correspondence

between surface points and Alpha Carbon (C_α) atoms is then established based on Euclidean distances to achieve precise feature alignment. Finally, a one-shot decoder is used to generate the predicted amino acid sequence. Experimental results demonstrate that SurfFold achieves state-of-the-art performance on the CATH4.2 dataset ([Orengo et al. 1997](#)). The model consistently achieves sequence-recovery rates of 62.47% in the test dataset. Additional experiments validate the critical role of the representation alignment module, and ablation studies confirm the necessity and effectiveness of each component within the framework. Furthermore, the homologous structure inverse folding experiment establishes SurfFold’s excellent performance in designing homologous proteins.

2 Materials and methods

In this section, we provide detailed implementation specifics of our proposed SurfFold. SurfFold integrates protein surface information with structural information to predict protein sequences. Additionally, we propose a representation alignment method to unify the protein surface representations and structural representations obtained from distinct modules. Thus, the implementation of SurfFold primarily consists of three modules: Structure Encoder Module, Surface Encoder Module, Representation Alignment Module. [Figure 1](#) depicts the SurfFold architecture, highlighting the interactions among the three modules.

2.1 Structure Encoder Module

In contrast to most inverse folding models, our Structure Encoder Module leverages both backbone and side-chain structures of proteins. Using atomic coordinates from PDB files, we extract the backbone and side-chain structure to construct the initial node features of the protein graph. Within the Structure Encoder Module, we construct a protein graph based on distances between C_α atoms, establishing an edge between two C_α nodes if their distance falls below a certain threshold. Subsequently, the combined backbone and side-chain features serve as initial node representations, which are processed by an EdgeGraphConv ([Wang et al. 2022](#)) network to generate residue-level structural representations. EdgeGraphConv is similar to GraphConv ([Morris et al. 2019](#)), with the only difference being that EdgeGraphConv applies the Hadamard product between node features and edge features.

2.1.1 Backbone structure features

Following the backbone structure feature extraction approach described in ProNet ([Wang et al. 2022](#)), we define a local coordinate system for each amino acid residue. Each amino acid residue is represented as a node in the graph, and we extract spherical coordinates ($d_{ji}, \theta_{ji}, \phi_{ji}$) of node j in the local coordinate system of node i , as well as the rotation angle ω_{ji} of the edge between these nodes, to serve as basic backbone structural features. Furthermore, to determine the relative orientation between two amino acids, we calculate three Euler angles ($\tau_{ji}^1, \tau_{ji}^2, \tau_{ji}^3$) between their respective backbone coordinate systems. These Euler angles represent the relative orientation between the two amino acid backbone coordinate systems and help to describe their spatial relationship. The angles and coordinate information are then combined to form the structural features of the protein backbone

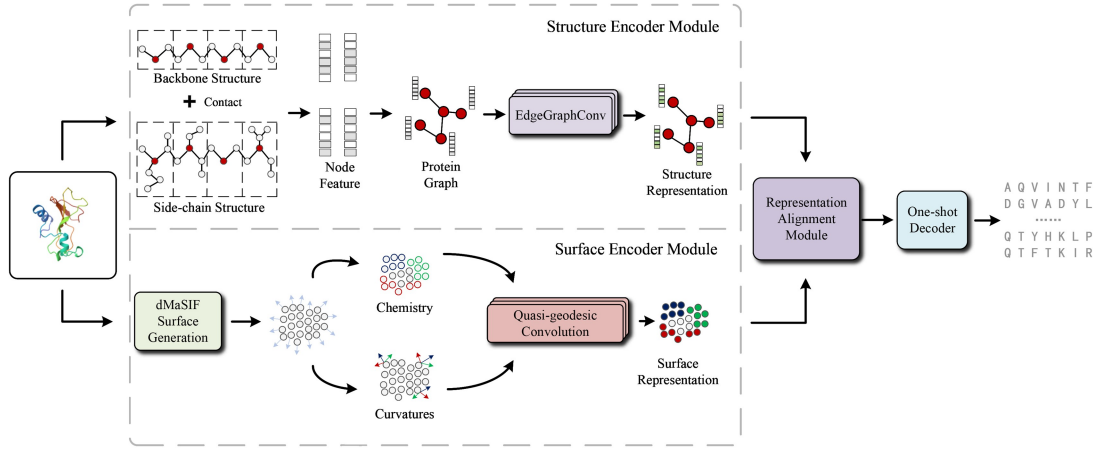


Figure 1. The overall framework of SurfFold. The Structure Encoder Module extracts representations from the protein structure. The Surface Encoder Module utilizes the dMaSIF surface generation method to generate the protein surface from its structure and then extracts surface representations. The Representation Fusion Module integrates representations from these two modalities to form a fused representation. This fused representation is ultimately decoded via one-shot decoding to predict the protein sequence.

$$Feature_b = (d_{ji}, \theta_{ji}, \phi_{ji}) \cup (\tau_{ji}^1, \tau_{ji}^2, \tau_{ji}^3) \quad (1)$$

where $Feature_b$ is the feature of protein backbone structure.

2.1.2 Side-chain structure features

To effectively integrate with protein surface information, we adopt the side-chain structural feature extraction method described in ProNet (Wang *et al.* 2022) during our data preprocessing stage. We specifically extract structural information from the protein side chains without incorporating detailed atomic information, thus avoiding potential data leakage issues. Based on known bond lengths and bond angles among atoms within side chains, we calculate four side-chain torsion angles ($\chi_1, \chi_2, \chi_3, \chi_4$). These torsion angles are subsequently integrated with backbone structural information to generate the initial node features for the protein graph

$$Feature_p = Feature_b \cup (\chi_1, \chi_2, \chi_3, \chi_4) \quad (2)$$

where $Feature_p$ is the node feature of protein graph.

2.2 Surface Encoder Module

To address critical information that protein internal structures fail to capture, SurfFold introduces a Surface Encoder Module to thoroughly exploit information from the protein surface. The Surface Encoder Module leverages the dMaSIF (Sverrisson *et al.* 2021) method, significantly enhancing both the efficiency and accuracy of surface generation compared to traditional methods such as MSMS (Ewing and Hermisson 2010) and MaSIF (Gainza *et al.* 2020). Initially, the module performs protein surface sampling, subsequently extracting chemical (such as electrostatic properties) and geometric attributes (such as surface curvature) from the generated surface point cloud. Finally, quasi-geodesic convolutions integrate and effectively learn the surface information, resulting in robust protein surface representations.

2.2.1 Surface generation

Protein surface reconstruction is directly based on atomic coordinates and types, avoiding explicit mesh preprocessing. Utilizing the Smooth Distance Function (SDF) (Sverrisson *et al.* 2021), we define the protein surface as an isosurface.

Initial points are generated via Gaussian sampling and projected onto the isosurface by gradient optimization. The resulting point cloud is then downsampled to yield a uniformly distributed protein surface point cloud. The SDF is defined as:

$$\sigma(x) = \frac{\sum_{k=1}^A \exp\left(-\frac{\|x - a_k\|}{\sigma_k}\right) \sigma_k}{\sum_{k=1}^A \exp\left(-\frac{\|x - a_k\|}{\sigma_k}\right)} \quad (3)$$

$$\text{SDF}(x) = -\sigma(x) \cdot \log\left(\sum_{k=1}^A \exp\left(-\frac{\|x - a_k\|}{\sigma_k}\right)\right) \quad (4)$$

where σ_k denotes the radius of atom a_k . Subsequently, the normals for each point on the surface are computed by normalizing the gradient of the SDF:

$$\hat{n}_i = \frac{\nabla \text{SDF}(x_i)}{\|\nabla \text{SDF}(x_i)\|} \quad (5)$$

where the ∇ denotes the Gradient operator. This point cloud-based generation method reduces preprocessing time and significantly improves computational efficiency for large-scale protein data.

2.2.2 Surface chemical features

During chemical feature extraction, our approach eschews traditional handcrafted descriptors, such as electrostatic potentials, instead employing deep networks to autonomously learn chemical environmental information. For each surface point, we identify its 16 nearest atoms. The atomic types (encoded via one-hot encoding) and the inverse distances to the surface points are used as input features. These features are first processed by a small MLP, and then passed through another MLP to yield the final chemical features. This end-to-end, data-driven strategy enables adaptive learning of diverse protein surface environments and effectively characterizes essential chemical properties, demonstrating superior performance in dMaSIF (Sverrisson *et al.* 2021) benchmarks.

2.2.3 Surface structure features

For structural feature extraction, the module emphasizes capturing local geometric information. Considering each surface point as a reference, we compute multi-scale curvature features (mean curvature and Gaussian curvature) from neighboring points. These curvatures effectively describe both detailed and overall surface shape at various scales. The structural features obtained are concatenated with corresponding chemical features to serve as inputs to subsequent convolutional layers. By adjusting Gaussian weight windows and quasi-geodesic neighborhood sizes, the model adeptly adapts to different local geometric patterns and maintains robustness against spatial rotations and translations, thereby improving the model’s ability to capture and characterize complex protein surface structures.

2.2.4 Quasi-geodesic convolution

After obtaining chemical and geometric surface features, quasi-geodesic convolutions capture local spatial structural features and map neighboring point features into a higher-dimensional space. The convolution operation selects neighboring points using Gaussian weighting, approximating geodesic distances by considering spatial distances and angles between normals. Neighboring point features are encoded within local coordinate systems and aggregated via a trainable MLP to form new point features. This approach is invariant to 3D spatial rotations and translations, while effectively exploiting complex surface spatial information.

2.3 Representation alignment module

The number of protein surface points generated by dMaSIF (Sverrisson *et al.* 2021) is substantially larger than the number of structural points in the protein backbone. After learning separate representations from the surface and structural modules, fusing these representations poses challenges in both geometric alignment and semantic consistency. To address this, we design a representation fusion module that integrates outputs from the two branches to facilitate downstream protein sequence prediction. The architecture of this module is illustrated in Fig. 2.

First, we use a KDTree to identify the eight nearest surface points for each C_α atom in the protein backbone, ensuring efficient nearest-neighbor search. The representations of these eight surface points are aggregated through average pooling to generate a surface representation aligned with each corresponding C_α atom. This operation ensures that the number of surface representations matches that of the structural nodes. Subsequently, we employ separate MLPs to project the structural and surface representations into a shared dimension space:

$$\begin{cases} \mathbf{a}' = W_a \mathbf{a} + b_a, \\ \mathbf{b}' = W_b \mathbf{b} + b_b \end{cases} \quad (6)$$

where $\mathbf{a}$ and $\mathbf{b}$ denote the structural and surface representations, respectively, and W_a , W_b , b_a , and b_b are learnable parameters.

To further integrate the protein structure and surface information, we apply a cross-attention mechanism to jointly refine both representations:

$$\begin{cases} \text{Attn}_A = \text{MHA}(\mathbf{a}', \mathbf{b}', \mathbf{b}'), \\ \text{Attn}_B = \text{MHA}(\mathbf{b}', \mathbf{a}', \mathbf{a}') \end{cases} \quad (7)$$

where $\text{MHA}(\cdot)$ denotes the multi-head attention operation, with the first argument as the query, and the second and third arguments as the key and value, respectively.

We then apply residual connections and layer normalization to the attention outputs, resulting in enhanced fused representations. Finally, the fused representation used for sequence prediction is generated by passing through an MLP:

$$\begin{aligned} \mathbf{a}_1 &= \text{LayerNorm}(\mathbf{a}' + \text{Dropout}(\text{Attn}_A)), \\ \mathbf{b}_1 &= \text{LayerNorm}(\mathbf{b}' + \text{Dropout}(\text{Attn}_B)), \\ \mathbf{f}_{\text{fused}} &= \text{MLP}_{\text{fusion}}([\mathbf{a}_1; \mathbf{b}_1]) \end{aligned} \quad (8)$$

where $[\cdot; \cdot]$ denotes concatenation along the feature dimension.

2.4 One-shot decoder

To generate the protein sequence from the fused representation, we adapt a one-shot decoder inspired by PiFold (Gao *et al.* 2023a). Unlike autoregressive decoders, which predict amino acids sequentially and often require beam search, the one-shot decoder predicts the entire sequence simultaneously, significantly improving computational efficiency. Despite its simplicity, PiFold (Gao *et al.* 2023a) demonstrates that this approach achieves competitive sequence-recovery accuracy. Following this design, we implement our decoder as a multi-layer perceptron (MLP) that maps the fused representation $\mathbf{f}_{\text{fused}}$ directly to a probability distribution over the amino acid vocabulary at each position. In addition, this approach reduces the risk of exposure bias commonly associated with autoregressive decoding, as each amino acid prediction is conditionally independent given the fused representation. This approach strikes a balance between decoding speed and prediction quality, making it ideal for large-scale inverse folding applications.

3 Results and Discussion

3.1 Baseline

To ensure a fair and comprehensive evaluation, we compare our proposed inverse folding model with several representative state-of-the-art methods. The selected baselines include both classical algorithms and advanced deep learning models. These consist of graph neural network-based methods such as GVP (Jing *et al.* 2020), GCA (Tan *et al.* 2022), and StructGNN (Tan *et al.* 2022), as well as recently developed neural frameworks like ProteinMPNN (Dauparas *et al.* 2022), LM-design (Zheng *et al.* 2023), PiFold (Gao *et al.* 2023a), Knowledge-Design (Gao *et al.* 2024a), and ScFold (Zhong *et al.* 2025). Additionally, we include SurfPro (Song *et al.* 2024) as a surface-based method and ESM-IF (Hsu *et al.* 2022) as a powerful pretrained model. Together, these baselines represent diverse algorithmic paradigms in the field of protein sequence design, facilitating a comprehensive and objective evaluation of our model’s performance and generalization capability. To ensure a fair comparison with existing methods, we adopt the same evaluation metrics and dataset as PiFold (Gao *et al.* 2023a). The experimental data are the average of results across different random seeds. Further

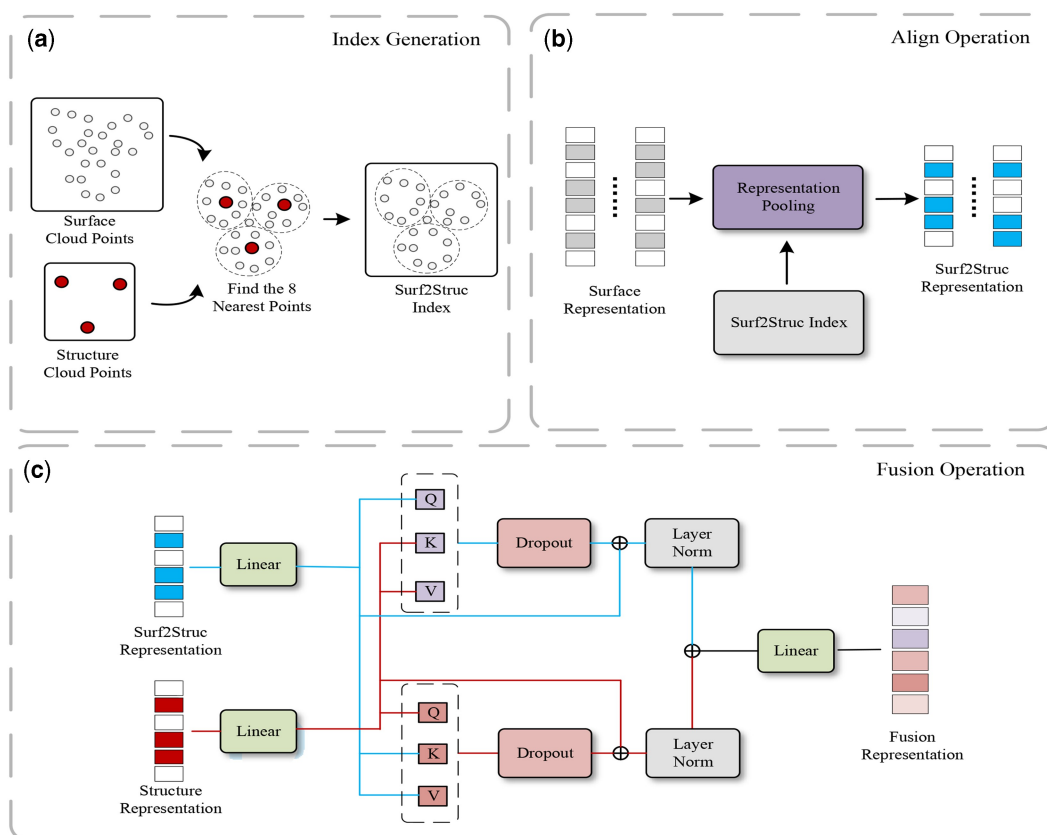


Figure 2. (a) The 8-nearest-neighbor alignment operation. During the data processing stage, the model first identifies the eight nearest surface point clouds for each structural point cloud based on their 3D coordinates. These indices are stored as Surf2Struc index. (b and c) Constitute the representation alignment module, where (b) represents the surface representation alignment operation. Using the precomputed Surf2Struc indices, this module processes the surface representation (from the Surface Encoder Module) to align them with the structure representation (from the Structure Encoder Module), thereby generating Surf2Struc representation with identical dimensions. (c) Refers to the representation fusion operation within the Representation Alignment Module. The model employs bidirectional cross-attention to fuse the structure representations and Surf2Struc representation into a unified Fusion representation, which is subsequently used for one-shot decoding to predict the protein sequence.

Table 1. CATH4.2 dataset experiments.

Model	Perplexity ↓			Recovery ↑		
	Short	Single	All	Short	Single	All
GVP	7.23	7.84	5.36	30.60	28.95	39.47
GCA	7.09	7.49	6.05	32.62	31.10	37.64
StructGNN	8.29	8.74	6.40	29.44	28.26	35.91
ProteinMPNN	6.21	6.68	4.61	36.35	34.43	45.96
LM-design	6.77	6.46	4.52	37.88	42.47	55.65
ESM-IF	8.18	6.33	6.44	31.3	38.5	38.3
PiFold	6.04	6.31	4.55	39.84	38.53	51.66
SurfPro	–	–	3.13	–	–	57.78
Knowledge-Design	5.48	5.1	3.46	44.66	45.45	60.77
ScFold	5.80	5.99	4.61	41.60	40.10	52.22
SurfFold(our)	3.20±0.06	3.32±0.03	2.95±0.01	56.18±1.27	54.38±0.34	62.05±0.54

Note: The best performance is highlighted in bold.

dataset descriptions and evaluation metrics are provided in the [Supplementary Materials](#), available as [supplementary data](#) at *Bioinformatics* online.

3.2 Result

The experimental results on CATH4.2 ([Orengo et al. 1997](#)), shown in [Table 1](#), demonstrate that SurfFold achieves the lowest perplexity across all test partitions, benefiting from the integration of 3D structural and protein surface

information. In terms of sequence recovery, SurfFold attains 56.18%, 54.38%, and 62.05% on short-chain, single-chain, and full-test partitions, respectively. While the gain over the second-best model is modest (1.3%) on the full-test set, SurfFold shows notable advantages on short-chain and single-chain partitions, suggesting that its effectiveness is largely independent of protein sequence length. Moreover, compared with SurfPro ([Song et al. 2024](#)), which uses only surface information, SurfFold improves sequence recovery by

Table 2. TS50 and TS500 dataset experiments.

Model	TS50		TS500	
	Perplexity	Recovery	Perplexity	Recovery
GVP	4.71	44.14	4.20	49.14
GCA	5.09	47.02	4.72	47.74
StructGNN	5.40	43.89	4.98	45.69
ProteinMPNN	3.93	54.43	3.53	58.08
PiFold	3.86	58.72	3.44	60.42
Knowledge-Design	3.10	62.79	2.86	69.19
ScFold	3.71	59.32	—	61.59
SurfFold(our)	2.90±0.05	64.16±0.27	2.74±0.02	69.32±0.40

Note: The best performance is highlighted in bold.

Table 3. Ablation experiments.

		SurfFold	Model1	Model2	Model3	Model4	Model5
Protein features	Backbone	✓	✓	✓	✓	✓	✓
	Side-chain	✓	✓			✓	✓
	Surface	✓			✓	✓	✓
Alignment Module	8-nearest alignment	✓			✓	✓	
	Cross-attention	✓			✓		✓
Results	Perplexity	2.95±0.01	3.23±0.02	13.27±0.36	8.20±0.06	3.24±0.02	3.12±0.03
	Recovery	62.05±0.54	56.94±0.62	18.31±1.52	36.03±0.96	57.61±0.53	60.04±1.21

4.27%, highlighting the value of combining structural and surface features for more accurate protein sequence prediction.

In Table 2, SurfFold achieves the lowest perplexity (2.90/2.74) and the highest sequence recovery (64.16%/69.32%). Compared with the prior best, Knowledge-Design, SurfFold improves recovery by 1.37 percentage points on TS50 and still leads by 0.13 points on TS500. Against strong baselines on TS50, it gains +9.73 and +5.44 points over ProteinMPNN and PiFold, respectively, indicating that fusing 3D structural and surface cues markedly enhances sequence design quality. Taken together, these consistent gains across small (TS50) and medium (TS500) benchmarks suggest strong generalization across dataset scales and protein diversity.

3.3 Ablation studies

We conducted a series of ablation studies to rigorously assess the contribution of each component in our model and understand their individual impacts on performance. Based on the architecture described in Section 2, we constructed several variant models. Regarding the protein modalities, we evaluated the following variants in addition to the full model: (1) a backbone-only variant, (2) a backbone with surface variant, and (3) a backbone with side-chain variant. At the module level, we also tested versions that omit the 8-nearest-neighbor aggregation for surface representation and those that exclude the bidirectional cross-attention mechanism for integrating surface and structural features. The performance of each ablated model is summarized in Table 3.

The ablation results reveal that side-chain information alone yields a substantial performance gain, yet it is insufficient to achieve state-of-the-art accuracy. Only when side-chain features are effectively fused with surface representations does the model realize its full potential. Moreover, these experiments demonstrate that our proposed 8-nearest-

neighbor aggregation and bidirectional cross-attention fusion significantly mitigate conflicts between different modal embeddings, thereby enhancing overall model performance.

3.4 Representation alignment analysis and visualization

To further assess the efficacy of our representation alignment module, we visualized representations immediately before and after its application. Using saved model parameters, we randomly sampled test instances and projected them into two dimensions using t-SNE (Fig. 3). In these representations, the raw surface representation appears as a diffuse, unstructured cloud, reflecting the fact that the surface point cloud vastly outnumbers the structural point cloud and that surface descriptors are much higher in dimension, both of which introduce considerable noise. After applying aggregation of the eight nearest neighbors, the surface representation shows better local coherence and more closely mirrors the structural representation, with irrelevant noise effectively reduced. Nonetheless, although both types of representations tend to form clusters, significant class overlap remains. In contrast, once structure and surface representation are fused using cross-attention, the fused representation reveals clusters that are clearly separated. These results confirm that our alignment module substantially improves the discriminative power of the feature space, boosting inter-class separability and reducing noise, which in turn enables the one-shot decoder to predict protein sequences with greater accuracy.

3.5 Homologous structure inverse folding experiment analysis

To better validate the role of the protein surface and side chains introduced by SurfFold, we designed the homologous structure inverse folding experiment. Homologous structure inverse folding refers to the computational inverse design of amino acid sequences capable of folding into spatial

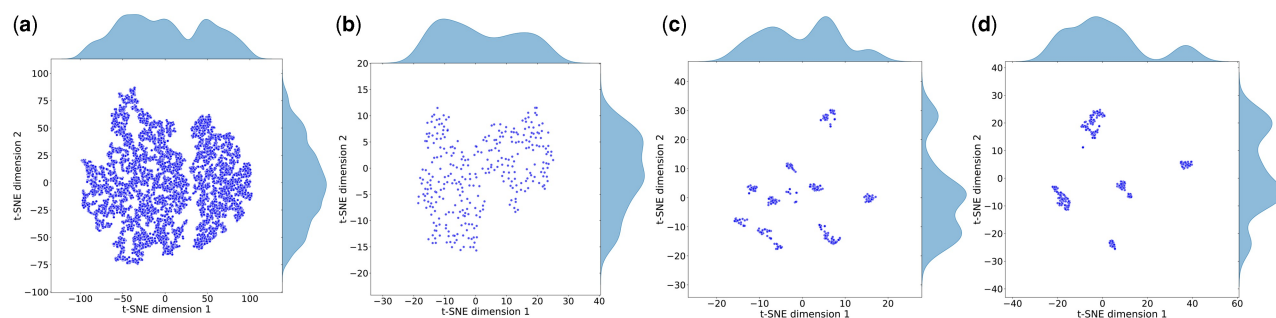


Figure 3. (a) The initial protein surface data. (b) The Surf2Struc representation after the 8-nearest-neighbor alignment. (c) The protein structural data. (d) The fusion representation after the representation alignment module.

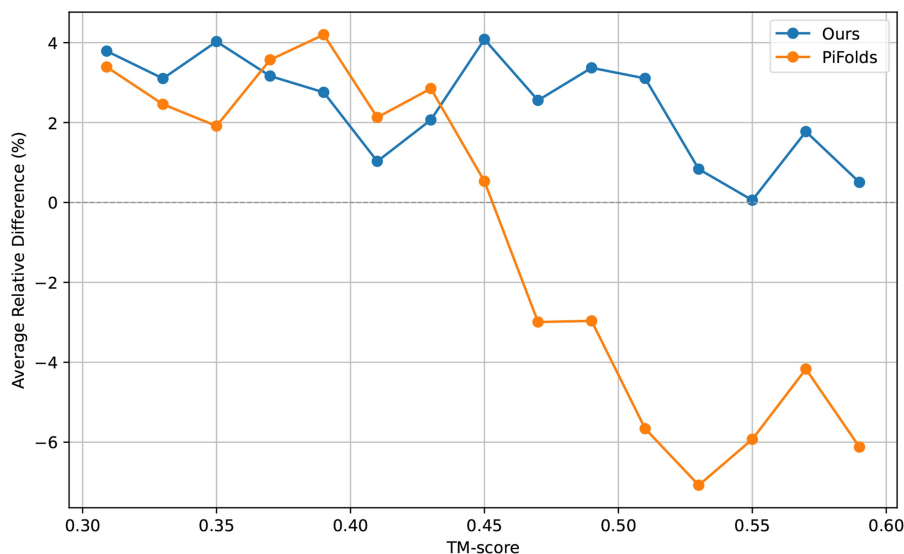


Figure 4. Relative difference between predicted and true protein sequence by TM-score.

conformations closely resembling a given homologous structural template, thereby enabling directed optimization or *de novo* functionalization of specific biological properties. We conducted pairwise structural TM-score analysis for all proteins in the test set, selecting pairs with a TM-score between 0.3 and 0.6. From these, we randomly sampled 1000 pairs based on TM-score distribution and grouped them into bins with a 0.02 interval. We then predicted the sequence of one protein in each pair using SurfFold and PiFold, comparing it to the native sequence of the other protein to measure sequence similarity. We also calculated the native sequence similarity between the protein pairs and computed the average relative difference between predicted and native similarity for each group. The average relative difference measures the relative error between the predicted sequence and the true sequence, defined as the difference in percent identity between the two, divided by the percent identity of the true sequence. The detail is shown in [Supplementary Materials](#), available as [supplementary data](#) at *Bioinformatics* online.

Figure 4 compares how the sequence prediction accuracy of the two protein inverse folding models [SurfFold and PiFold (Gao *et al.* 2023a)] varies with TM-score (structural similarity). The performance of SurfFold and PiFold (Gao *et al.* 2023a) is comparable when the TM-score is between 0.3 and 0.45. However, when the TM-score exceeds 0.45, the sequence similarity predicted by SurfFold becomes significantly higher than that predicted by PiFold (Gao *et al.*

2023a). Backbone-only geometry is insufficient for highly similar protein structures because it ignores surface chemical cues, causing key-residue mismatches and lowering accuracy. By jointly modeling surface geometry and chemistry (electrostatics, hydrophobicity), SurfFold captures constraints at functional sites—e.g., assigning Arg/Lys to negatively charged patches—and improves prediction. Even in medium-homology cases, conserved local physicochemical microenvironments guide SurfFold to place property-matched residues (e.g. Phe/Tyr in hydrophobic pockets), strengthening homology-based prediction and design.

4 Conclusion

This article presents a novel protein sequence prediction framework, SurfFold, which enhances sequence prediction accuracy by integrating protein structural information and surface information. SurfFold combines backbone, side-chain structures, and protein surface data, utilizing E(3)-equivariant neural networks and geodesic convolution networks to extract structural and surface features, respectively. We introduce a representation-alignment module to reconcile structural and surface embeddings, which markedly improves predictive performance. On CATH4.2 (Orengo *et al.* 1997), SurfFold surpasses prior models, with especially strong sequence-recovery gains on short- and single-chain proteins; alignment visualizations and homologous inverse folding

results further confirm the benefit of jointly modeling structural and surface cues. Currently limited to proteins, the framework has not yet been tested on nucleic acids or protein–nucleic acid complexes; future work will extend SurfFold to these settings.

Author contributions

Darong Li (Conceptualization [equal], Investigation [lead], Methodology [lead], Writing—original draft [lead], Writing—review & editing [equal]), Lian Shen (Conceptualization [lead], Methodology [equal], Writing—original draft [equal], Writing—review & editing [equal]), Meijia Song (Formal analysis [equal], Resources [equal]), Deyi Li (Formal analysis [equal], Resources [equal]), Juan Liu (Project administration [supporting], Supervision [supporting]), and Xiangrong Liu (Project administration [lead], Supervision [lead])

Supplementary data

[Supplementary data](#) is available at *Bioinformatics* online.

Conflict of interest: The authors declare no conflict of interest.

Funding

This work was supported by the National Key Research and Development Program of China (2023YFF1205600, 2024YFF1206204), National Natural Science Foundation of China (grant nos. 62372391, 62072384), the Fujian Provincial Major Science and Technology Project (2022YZ040011), the Wang Deyao Outstanding Graduate Scholarship Program of Xiamen University, the Scientific Research Foundation of State Key Laboratory of Vaccines for Infectious Diseases, Xiang An Biomedicine Laboratory (2023XAKJ0102045).

Data availability

The source code and data are available at <https://github.com/jiudizhengf/SurfFold>

References

- Bertoni M, Kiefer F, Biasini M *et al.* Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* 2017;7:10480.
- Dauparas J, Anishchenko I, Bennett N *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 2022;378:49–56.
- Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science* 2012;338:1042–6.
- Ewing G, Hermisson J. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 2010;26:2064–5.
- Fang Y, Jiang Y, Wei L *et al.* DeepProSite: structure-aware protein binding site prediction using ESMFold and pretrained language model. *Bioinformatics* 2023;39:btad718.
- Gainza P, Sverrisson F, Monti F *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* 2020;17:184–92.
- Gao Z, Tan C, Li SZ. Alphadesign: a graph protein design method and benchmark on alphafold. *arXiv*, arXiv:2202.01079, 2022, preprint: not peer reviewed.
- Gao Z, Tan C, Li SZ. PiFold: Toward effective and efficient protein inverse folding. *International Conference on Learning Representations (ICLR)*, 2023a. <https://openreview.net/forum?id=oMsN9TYwJ0j>.
- Gao Z, Tan C, Chen X *et al.* KW-Design: Pushing the limit of protein design via knowledge refinement. *The Twelfth International Conference on Learning Representations (ICLR)*. 2024a. <https://openreview.net/forum?id=mpqMVWggjn>.
- Gao Z, Tan C, Zhang Y *et al.* ProteinInvBench: benchmarking protein inverse folding on diverse tasks, models, and metrics. In: *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2023b, 68207–20.
- Gao Z, Wang J, Tan C *et al.* UniIF: unified molecule inverse folding. In: *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2024b, 135843–60.
- Hsu C, Verkuil R, Liu J *et al.* Learning inverse folding from millions of predicted structures. In: *Proceedings of the 39th International Conference on Machine Learning*, Cambridge, MA, USA: PMLR, 2022, 8946–70.
- Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature* 2016;537:320–7.
- Ingraham J, Garg V, Barzilay R *et al.* Generative models for graph-based protein design. In: *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2019.
- Jing B, Eismann S, Suriana P *et al.* (2021). Learning from protein structure with geometric vector perceptrons. *International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=1YLJDvSx6J4>.
- Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with alphafold. *Nature* 2021;596:583–9.
- Li P, Liu Z-P. GeoBind: segmentation of nucleic acid binding interface on protein surface with geometric deep learning. *Nucleic Acids Res* 2023;51:e60.
- Lin Z, Akin H, Rao R *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;379:1123–30. <https://doi.org/10.1126/science.ade2574>.
- Morris C, Ritzert M, Fey M *et al.* Weisfeiler and Leman go neural: higher-order graph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, Vol. 33. 2019, 4602–9.
- Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545–600.
- Orengo CA, Michie AD, Jones S *et al.* CATH – a hierarchical classification of protein domain structures. *Structure* 1997;5:1093–108.
- Roche R, Moussad B, Shuvo MH *et al.* EquiPNAS: improved protein–nucleic acid binding site prediction using protein–language-model-informed equivariant deep graph neural networks. *Nucleic Acids Res* 2024;52:e27.
- Rose GD. Ramachandran maps for side chains in globular proteins. *Proteins: Struct Funct Bioinform* 2019;87:357–64.
- Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. In: *Proceedings of the 38th International Conference on Machine Learning*, Cambridge, MA, USA: PMLR, Vol. 139. 2021, 9323–32.
- Senior AW, Evans R, Jumper J *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* 2020;577:706–10.
- Song Z, Huang T, Li L *et al.* Surfpro: functional protein design based on continuous surface. In: *International Conference on Machine Learning*, Cambridge, MA, USA: PMLR, 2024, 46074–88.
- Sverrisson F, Feydy J, Correia BE *et al.* Fast end-to-end learning on protein surfaces. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021. New York, NY, USA: IEEE, 15272–81.

- Tan C, Gao Z, Xia J *et al.* Global-context aware generative protein design. *ICASSP 2023 – IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2023;1–5. <https://doi.org/10.1109/ICASSP49357.2023>
- Wang L, Liu H, Liu Y *et al.* Learning hierarchical protein representations via complete 3D graph networks. *International Conference on Learning Representations (ICLR)* 2023.
- Xu S, Shen L, Zhang M *et al.* Surface-based multimodal protein–ligand binding affinity prediction. *Bioinformatics* 2024;40:btae413.
- Yang KK, Zanichelli N, Yeh H. Masked inverse folding with sequence transfer for protein representation learning. *Protein Eng Des Sel* 2023;36:gzad015.
- Yuan M, Shen A, Ma Y *et al.* ProteinF3S: boosting enzyme function prediction by fusing protein sequence, structure, and surface. *Brief Bioinform* 2024;26:bbae695.
- Zheng Z, Deng Y, Xue D *et al.* Structure-informed language models are protein designers. In: *Proceedings of the 40th International Conference on Machine Learning*, Cambridge, MA, USA: PMLR, 2023, 42317–38.
- Zhong J, Zou Z, Qiu J *et al.* ScFold: a GNN-based model for efficient inverse folding of short-chain proteins via spatial reduction. *Brief Bioinform* 2025;26:bbaf156.
- Zhu X, Lopes PE, Shim J *et al.* Intrinsic energy landscapes of amino acid side-chains. *J Chem Inf Model* 2012;52:1559–72.