SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2026, Vol. 69, Iss. 11, 210209:1–210209:13
https://doi.org/10.1007/s11432-026-5239-1

Special Topic: Intelligent Characterization and Modeling for Complex Systems

OPUS-Design: a 3D convolutional approach for protein sequence design targeting complex spatial microenvironments

Gang XU[1], Qinghua WANG[2] & Jianpeng MA[1,3*]

[1]*Multiscale Research Institute of Complex Systems, Fudan University, Shanghai 200433, China*
[2]*Center for Biomolecular Innovation, Harcam Biomedicines, Shanghai 200131, China*
[3]*School of Physics and Electronics, Hunan Normal University, Changsha 410082, China*

Received 19 March 2026/Revised 6 October 2026/Accepted 29 October 2026/Published online 10 November 2026

Abstract Complex three-dimensional (3D) protein environments, including monomers, oligomers, and ligand-bound small-molecule complexes, form a canonical class of biological systems. Protein sequence design based on backbone structure, commonly referred to as protein inverse folding, is central to modeling these intricate systems, as it seeks to uncover the intrinsic link between 3D structural topology and target amino acid sequences within complex spatial microenvironments. State-of-the-art methods for this task are dominated by graph-based models such as ProteinMPNN and ESM-IF, which deliver strong performance. However, 3D convolutional neural networks (3DCNNs) offer a promising alternative. These networks feature inherent rotation invariance and excel at capturing dense local spatial contexts in complex protein microenvironments, yet their full potential for protein inverse folding remains underexplored. Here, we introduce OPUS-Design, a two-stage 3DCNN-based framework for protein sequence design. Tested on CASP15 and CAMEO targets spanning monomers, oligomers, and ligand-bound complexes, OPUS-Design matches or outperforms top graph-based methods in sequence recovery. These findings validate the efficacy of 3DCNN for protein inverse folding and suggest its potential for broader applications in related structural prediction tasks.

Keywords protein sequence design, protein inverse folding, protein complex, protein spatial microenvironments, 3D convolutional neural networks

Citation Xu G, Wang Q H, Ma J P. OPUS-Design: a 3D convolutional approach for protein sequence design targeting complex spatial microenvironments. Sci China Inf Sci, 2026, 69(11): 210209, https://doi.org/10.1007/s11432-026-5239-1

1 Introduction

Complex systems, including biological macromolecule networks, cell systems, and natural ecosystems, form the foundation of modern cross-disciplinary scientific research. As a typical biological complex system, proteins fold into complex three-dimensional (3D) spatial structures across diverse functional states, such as homomeric and heteromeric oligomers as well as ligand-bound complexes. These structures feature detailed spatial layouts, multi-level atomic and domain interactions, and a close connection between structure and function. Within this context, rational protein design has emerged as a powerful tool in many applications, such as therapeutic engineering [1,2], binder development [3–5], biosensor construction [6,7], nanomaterial synthesis [8], and enzyme optimization [9–12]. Recent advances also include taxonomic-guided diffusion models, such as TaxDiff, for controllable protein sequence generation [13]. One critical sub-problem in this area is protein sequence design based on backbone, also known as protein inverse folding, where the objective is to deduce the optimal amino acid sequence from a predefined backbone structure, characterized by the specific positions of backbone atoms N, C_α, C, and O. To address this problem, over the past few years, many methods have been developed.

Given the significant computational costs associated with traditional physics-based approaches, which often result in unsatisfactory accuracy [14], the research community has been increasingly focusing on developing deep learning models as a viable solution in the field. For example, SPIN [15] and SPIN2 [16] are two pioneering studies developed by Zhou's group; they employ several fully-connected layers as their network architectures. Meanwhile, some researchers adopt a 3D convolutional neural network (3DCNN) to extract distinctive features for each residue [17,18]. These methods typically require a separate preprocessing step and perform independent predictions for each residue, such that the prediction of one residue does not influence or interact with others.

* Corresponding author (email: jpma@fudan.edu.cn)

© Science China Press 2026

Recently, numerous graph-based models have emerged and have demonstrated their potential in achieving superior sequence recovery rates. For example, PiFold [19] incorporates a unique residue feature extractor alongside PiGNN layers, enabling the one-shot generation of protein sequences with enhanced recovery rates. ProteinMPNN [20] and ESM-IF [21] leverage graph neural network (GNN)-based modules to encode backbone structure and subsequently decode the corresponding sequences in an end-to-end autoregressive fashion. It is noteworthy that ESM-IF significantly augments the training data by approximately three orders of magnitude through the utilization of AlphaFold2 [22] to predict structures for 12 million protein sequences. Consequently, ESM-IF achieves a marked improvement over others. Furthermore, some researchers have incorporated protein evolutionary features derived from protein language models, such as ESM-2 [23], to further enhance the accuracy of their models [24, 25].

Recent developments in convolutional neural networks include low-rank Winograd transformations for efficient 3D convolution [26] and semi-tensor product-based formulations for irregular and cross-dimensional data [27]. Convolutional architectures also integrate frequency-aware attention for physiological signal analysis, as demonstrated by CFAN for electrocardiogram (ECG)-based emotion recognition [28]. While graph-based architectures have become the dominant paradigm for protein inverse folding, yielding remarkable advances in sequence recovery accuracy, they are not the only viable strategy for encoding three-dimensional structural information. In contrast, 3DCNN offers inherent advantages in capturing dense spatial contexts, local atomic interactions, and rotation-invariant structural patterns directly from volumetric representations. Such a grid-based framework naturally models continuous structural environments and avoids the reliance on predefined graph topologies, making it highly suitable for characterizing complex and heterogeneous molecular structures. Despite these appealing properties, the potential of 3DCNN for backbone-guided protein sequence design remains largely underexplored.

In this study, we present OPUS-Design, a two-stage protein sequence design framework. The first stage adopts a 3DCNN-based feature extraction module adapted from our previous work OPUS-Rota5 [29], in which each atom is represented as a Gaussian sphere discretized into 3D voxels to capture fine-grained local environmental features of each residue. This module also incorporates constraints derived from side-chain conformations. In contrast to independent residue prediction, our framework enables the alteration of one residue to dynamically influence the local environment of neighboring residues, supporting a dynamically updated structural context during sequence design. In the second stage, we use a transformer-based feature aggregation module to integrate various types of information, which include 1D protein backbone features, 2D residue contact features, and 3D local environmental features. Moreover, the protein language model, ESM-2, is incorporated for further refinement. The structural representations and prediction strategies of OPUS-Design and related inverse-folding methods are summarized in Table A1 in Appendix A.

We evaluate the performance of OPUS-Design against state-of-the-art graph-based protein sequence design methods on recent targets from continuous automated model evaluation (CAMEO) and the 15th critical assessment of protein structure prediction (CASP15). The results demonstrate that OPUS-Design achieves comparable or even superior performance across monomers, oligomers, and ligand-bound complexes. Specifically, on the CAMEO monomer test set containing 339 targets, OPUS-Design achieves sequence recovery rates of 60.64% for all residues and 73.15% for core residues. In comparison, the second-best method ABACUS-T, a GNN and ESM-2-based approach, yields 56.18% and 69.12% for the same metrics, respectively. These findings underscore the reliability and versatility of 3DCNN in protein inverse folding, and imply that such volumetric-based deep learning strategies can serve as a powerful and generalizable paradigm for a wide spectrum of protein structure-related modeling tasks.

2 Method

2.1 Framework of OPUS-Design

The architecture of OPUS-Design is methodically organized into two distinct modules, a 3DCNN-based feature extraction module and a transformer-based feature aggregation module.

In the first feature extraction module, we employ an adapted version of the 3D U-Net from OPUS-Rota5 [29] to capture the local environmental and side-chain features of each residue. Inspired by DLPacker [30], we isolate a 20 Å box around each residue, encompassing its immediate surroundings, and subsequently align the atoms within this box to a standardized reference frame. To represent the atomic distribution within these isolated boxes, we employ a 3D Gaussian density kernel expansion, mapping each atom onto a $40 \times 40 \times 40$ voxel grid. We use 27 channels for representation: five dedicated to elemental identities (C, N, O, S, and others), one to encode partial charge information derived from the Amber99sb force [31], and 21 to distinguish between residue types (20 channels for common residues and one for others). The outputs of the first module include three branches. The first branch,

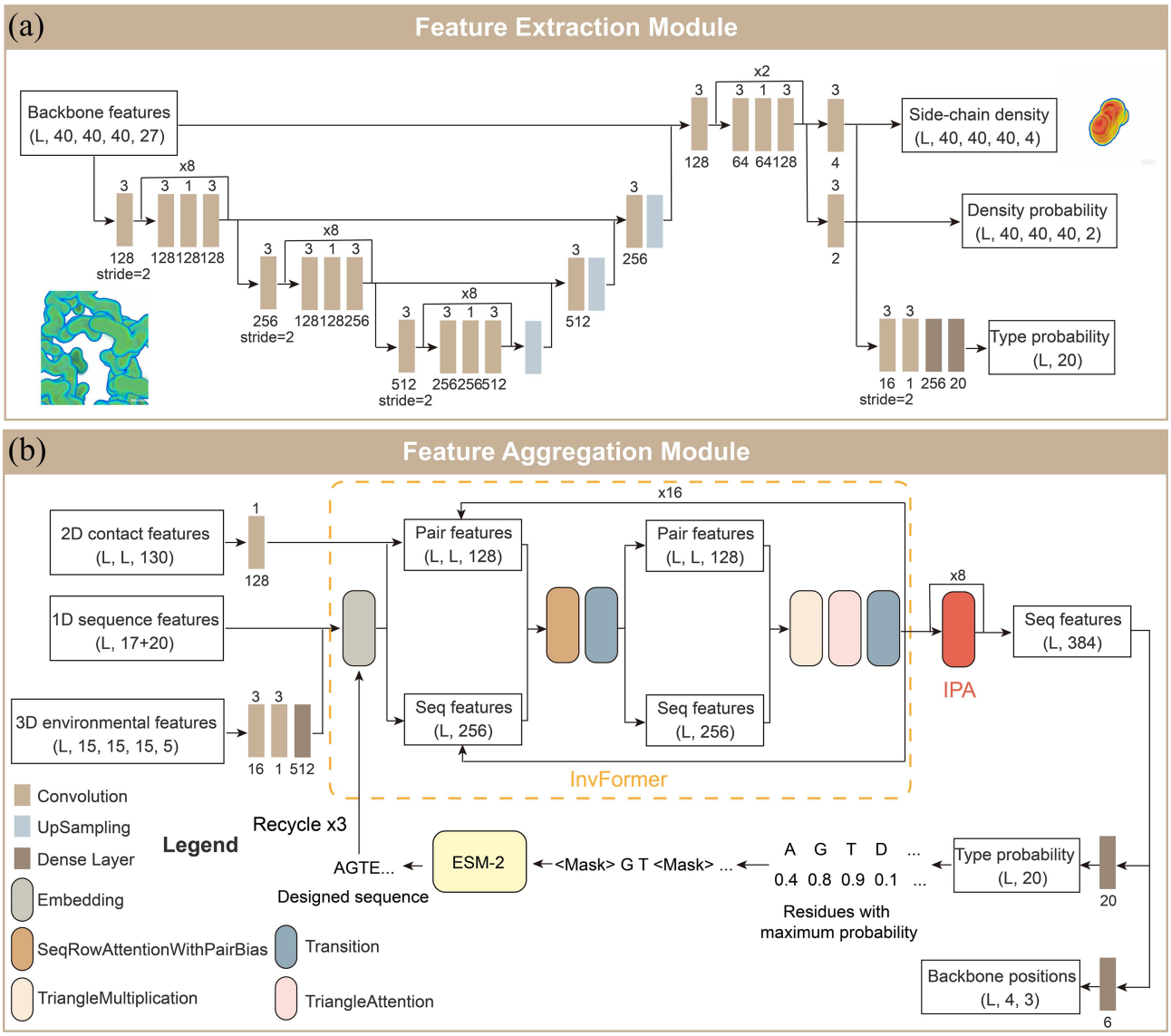

Figure 1 (Color online) Overview of the OPUS-Design workflow. (a) The first stage employs a 3DCNN for feature extraction, specifically designed to capture the 3D local environmental features of each residue. This module enables predictions to be made solely based on the local structural context. (b) The second stage adopts a transformer-based feature aggregation module to integrate diverse feature types: 1D protein sequence features, including the predicted probability distributions of each type of residue obtained from the first stage, 2D residue-residue backbone contact features, and the aforementioned 3D local environmental features. Subsequently, 8 IPA modules are deployed to reconstruct the 3D coordinates of backbone atoms. Furthermore, the workflow incorporates a refinement step utilizing ESM-2, where positions with a maximum predicted probability below 0.5 are designated as "mask" encodings, thereby undergoing further optimization.

comprising four channels (C, N, O, and S), depicts the 3D side-chain density of the target residue. The second branch, with two channels, quantifies the likelihood of each voxel within the grid containing side-chain density. Lastly, the third branch, containing 20 channels, offers the probability of each of the common residue types in the box. A detailed illustration of the first module is provided in Figure 1(a).

At inference, we remove protein side-chain atoms before constructing the local boxes, while retaining the backbone and relevant nonprotein atoms as structural context. The module processes residues sequentially in three passes. For each residue, it reconstructs the local box from the current structure and assigns the 3DCNN-predicted amino acid when the maximum residue-type probability is at least 0.5. An updated residue identity changes the residue-type and partial-charge features in subsequently constructed neighboring boxes, but does not add side-chain atoms or alter backbone coordinates. Between passes, ESM-2 evaluates the partially predicted sequence with unresolved positions masked and supplies predictions for the next pass. In the final pass, when the maximum 3DCNN probability is below 0.5, we use the ESM-2 probability distribution.

During the training process of the first module, we adopt the mean absolute error (MAE) loss between the

predicted and actual side-chain densities for the first branch, and the cross-entropy loss for the second and the third branches. Adam optimizer [32] is used for optimization with an initial learning rate of 1e−3. To mitigate the overfitting problem, we implement a learning rate scheduling strategy. Specifically, if the accuracy on the validation set declines following the completion of an epoch, we halve the learning rate. Training ceases after the rate has been reduced four times. We adopt an ensemble learning approach, training five independent models using the aforementioned configuration. The final prediction is determined as the average of their outputs. Each model is trained on four NVIDIA Tesla V100S GPUs.

In the second feature aggregation module, we integrate three distinct feature sets utilizing a transformer-based InvFormer module. In addition, we incorporate the invariant point attention (IPA) module, which is derived from AlphaFold2 [22], to introduce structural constraints. The three feature sets are described as follows. (1) The 1D protein sequence features, which encompass 3-state and 8-state secondary structures, alongside six backbone torsion angle features consisting of sine and cosine values for ϕ, ψ, and ω angles. (2) The 2D residue-residue backbone contact features are sourced from trRosetta [33] and include the C_β–C_β distance distributions and the orientational distributions of three dihedrals (ω, θ_{ab}, θ_{ba}) and two angles (ϕ_{ab}, ϕ_{ba}) between residues a and b. The distance distributions span from 2 to 20 Å, segmented into 36 bins at 0.5 Å intervals, with an additional bin accounting for distances exceeding 20 Å. The ϕ angle ranges from 0° to 180°, segmented into 18 bins at 10° intervals, inclusive of a non-contact bin. Both ω and θ vary from −180° to 180°, segmented into 36 bins at 10° intervals, also with a non-contact bin. Here, the positions of C_β atoms are determined utilizing the coordinates of backbone atoms. (3) The 3D local environmental features come from the first feature extraction module, comprising 5 channels. The first channel contains the local information of each residue, obtained by computing the density of backbone atoms in the corresponding box. The remaining 4 channels are the predicted 3D side-chain density of the target residue, obtained from the first output branch of the feature extraction module. For higher efficiency, we exclude the peripheral bins (specifically, 5 bins from each direction) and down-sample the density map from (40, 40, 40, 5) to (15, 15, 15, 5). Furthermore, the probability distribution for each residue type from the third output branch of the feature extraction module is also included.

The 3DCNN feature extraction module produces three outputs: side-chain density maps for four atom types (C, N, O, and S), an atom-type-independent side-chain density probability map, and a probability vector over the 20 standard amino acids. For the feature aggregation module, we retain the four atom-type-specific density maps because they preserve both spatial and atom-type information. The atom-type-independent output contains overlapping information and is therefore not passed separately. The four retained maps are combined with a backbone density map from the corresponding input box to form a five-channel 3D representation. The residue-type probabilities are retained separately as a 20-dimensional vector for each residue and concatenated with the other 1D residue features supplied to the feature aggregation module.

As shown in Figure 1(b), the InvFormer module is utilized to integrate both sequence and pair features. To construct the input for sequence features, the 3D local environmental features are initially processed through two 3D convolutional layers, subsequently flattened, and concatenated with 1D protein sequence features and probability distributions derived from the first feature extraction module. Concurrently, the 2D residue-residue backbone contact features undergo a 2D convolutional layer to formulate the input for pair features. In OPUS-Design, 16 InvFormer modules are deployed to enhance feature extraction capabilities. After feature aggregation, 8 IPA modules are employed to reconstruct the 3D coordinates of backbone atoms, enabling sequence features to encapsulate the structural information. The outputs of the feature aggregation module bifurcate into two branches: one dedicated to predicting the probabilities of 20 common residue types, and the other focused on predicting backbone conformations, utilizing the frame aligned point error (FAPE) loss from AlphaFold2 as guidance. Notably, the probability distributions for each residue type, along with the sequence and pair features from the preceding iteration, are reintegrated into the initial inputs, and this entire process is reiterated three times.

For multimeric targets, each residue-centered 20 Å box includes atoms from neighboring chains whenever they fall within its spatial boundaries, allowing the 3DCNN feature extraction module to capture interfacial environments. For feature aggregation, residues from all chains are concatenated into a single ordered list while retaining their original coordinates. The 2D pairwise geometric features are computed and encoded identically for intra-chain and inter-chain residue pairs. InvFormer uses these geometric features without an explicit chain-identity or intra-chain/inter-chain pair indicator; information about inter-chain interactions is provided through the geometric features rather than a separate chain-specific encoding.

Same as the training approach of the feature extraction module, we utilize the Adam optimizer with a starting learning rate of 1e−4. Furthermore, we train three distinct models and determine the final prediction by calculating the average of their respective outputs.

Additionally, we introduce the protein language model, ESM-2 [23], to deliver refined results from an evolutionary

standpoint. The protein language model is integrated into both feature extraction and aggregation modules. Briefly speaking, we designate the predicted residues with low confidence as "mask" and proceed to replace them with the outputs generated by ESM-2. Specifically, during feature extraction, unresolved positions are masked, while assigned residues provide unmasked context. ESM-2 can update assigned residues between 3DCNN passes. During feature aggregation, positions with a maximum predicted probability below 0.5 are masked, while higher-confidence predictions remain unchanged during ESM-2 refinement. For 3DCNN predictions and after each aggregation cycle, we use the ESM-2 probability distribution only when the structural-model confidence is below 0.5. Otherwise, we retain the structural-model distribution.

2.2 Datasets

In OPUS-Design, we employ the same training dataset as utilized by trRosetta [33], consisting of 15051 protein targets released before May 2018. To evaluate the performance across various methods, we adopt two distinct test sets comprising monomer targets: CAMEO and CASP15. CAMEO contains 339 monomer targets, sourced from the CAMEO website [34], spanning the recent five years, of which 65 were released between May 2021 and October 2021, 81 were released between April 2022 and June 2022, 82 were released between May 2023 and August 2023, 60 were released between March 2024 and June 2024, and 51 were released between January 2025 and April 2025. The CASP15 set, available for download from the CASP website[1)], comprises 44 additional monomer targets. To facilitate structural evaluation, we have constructed a test set named CAMEO-TM90, utilizing the targets sourced from the aforementioned CAMEO dataset. This test set comprises 164 monomer targets, each of which exhibits a TM-score exceeding 0.9 when comparing their native structures to the respective predicted structures obtained through ESMFold [23]. Meanwhile, we have gathered a test set for oligomer targets, CAMEO-oligomer, which comprises 133 oligomer structures. Among these, 50 targets were released by CAMEO between March and June 2024, and 83 were released between January 2025 and April 2025. Furthermore, we also used the PXM-2025 dataset collected by Protenix [35], which includes the ligand-bound protein complexes released in 2025. After filtering out entries with discontinuous backbones, 134 high-quality complexes remained.

2.3 Performance metrics

In this study, we adopt the sequence recovery rate as the primary quantitative metric, specifically quantifying the proportion of correctly predicted residues between the designed and native protein sequences. A high recovery rate signifies a close similarity between the designed and native sequences. Furthermore, we utilize the self-consistent TM-score (sc-TM), a metric that measures the structural similarity between predicted structures generated using the native sequence and the designed sequence. For the purposes of structural evaluation, we have limited our analysis to targets that demonstrate a TM-score exceeding 0.9 when comparing the native structure to the prediction obtained from ESMFold. This threshold is essential to ensure that we are comparing against reliable structural predictions, as an elevated structural similarity towards unreliable structures may ultimately yield misleading conclusions.

AA JSD (base-2 Jensen-Shannon divergence) [36] measures the difference between designed and native amino acid distributions, while mean BLOSUM62 [37] scores measure substitution similarity at corresponding positions. Per-residue charge and hydropathy MAE are calculated from the absolute differences between designed and native residues at each position, using estimated side-chain charge at pH 7 and the Kyte-Doolittle scale [38], respectively. ESM-C [39] mean pseudo-log-likelihood (PLL) is calculated with ESM-C 600M by masking one position at a time and averaging the log-probability of the observed residue over the evaluated positions.

In this study, in accordance with the FASPR [40], a residue is defined as a core residue when there are more than 20 residues within a 10 Å C_α–C_α distance threshold. Meanwhile, we classify residues as interfacial residues if they possess at least one neighboring residue, located on a distinct peptide chain, with a C_α–C_α distance of less than 8 Å. Similarly, we define a protein residue as ligand-related if the minimum distance between the C_α atom of that residue and any ligand atom is less than 8 Å.

We evaluate all methods on the same target lists and map their predictions to native residue positions before calculating sequence recovery. For oligomeric targets, the evaluated PiFold and ABACUS-T workflows combine all protein chains into a single-chain representation with sequential residue indices while preserving the backbone coordinates. ProteinMPNN, ESM-IF, LigandMPNN, and CARBonAra retain multichain structural information. For ligand-bound targets, PiFold, ProteinMPNN, and ESM-IF use protein structural information without ligand atoms, whereas LigandMPNN, CARBonAra, and ABACUS-T can incorporate ligand atoms according to their respective preprocessing rules. We use each method's default settings, except for ESM-IF, for which we set the

1) http://predictioncenter.org.

Table 1 Performance comparison of different methods on the CAMEO and CASP15 datasets, measured by all residues.

Method	Sequence recovery (%) ↑	Charge MAE ↓	Hydropathy MAE ↓	AA JSD ↓	BLOSUM62 ↑	ESM-C PLL ↑
CAMEO						
PiFold	50.30	0.29	1.52	0.05	2.35	−2.20
ProteinMPNN	46.49	0.30	1.61	0.08	2.10	−1.84
LigandMPNN	49.79	0.28	1.53	0.06	2.29	−1.90
CARBonAra	51.76	0.29	1.34	0.05	2.53	−2.17
ESM-IF	55.60	0.28	1.32	0.08	2.72	−1.67
ABACUS-T	56.18	0.26	1.30	0.07	2.66	−1.56
OPUS-Design	60.64	0.23	1.23	0.05	2.89	−1.76
CASP15						
PiFold	43.12	0.30	1.76	0.06	1.91	−2.35
ProteinMPNN	39.62	0.29	1.92	0.10	1.66	−1.91
LigandMPNN	42.68	0.28	1.87	0.09	1.79	−1.95
CARBonAra	46.98	0.29	1.49	0.06	2.26	−2.32
ESM-IF	48.28	0.28	1.65	0.12	2.19	−1.76
ABACUS-T	44.93	0.27	1.60	0.10	2.04	−1.67
OPUS-Design	50.69	0.25	1.44	0.06	2.39	−1.90

sampling temperature to 1e−6, as recommended in its instructions, to obtain more deterministic predictions. Note that some comparisons in our study may not be entirely fair to certain methods. Therefore, we do not claim that our 3DCNN-based method outperforms other method types in this scenario; rather, we aim to demonstrate the capability of 3DCNN methods in modeling complex spatial microenvironments.

Meanwhile, the evaluated models are not trained on a common dataset. OPUS-Design uses 15051 protein targets from the trRosetta training set, released before May 2018, together with pretrained ESM-2 for refinement. The baseline models have different training sources: PiFold uses CATH 4.2; ProteinMPNN uses protein data bank (PDB) biological assemblies available by August 2021; ESM-IF uses CATH structures and approximately 12 million predicted structures; LigandMPNN uses PDB assemblies available by December 2022; CARBonAra uses PDB biological assemblies; and ABACUS-T uses PDB structures released before April 2022 together with pretrained ESM sequence embeddings.

2.4 Data and software availability

The code and pre-trained models of OPUS-Design as well as the test sets used in the study can be downloaded[2)]. They are freely available for academic use.

3 Results

3.1 Performance of different methods on monomer targets

We evaluate the performance of OPUS-Design against several leading methods, namely PiFold [19], ProteinMPNN [20], LigandMPNN [41], ESM-IF [21], CARBonAra [42], and ABACUS-T [25], utilizing two monomer test sets: CAMEO and CASP15. The results show that OPUS-Design outperforms the competing methods in terms of the sequence recovery rate, both when considering all residues (Table 1) and focusing specifically on core residues (Table A2). It is worth noting that the training data employed by ESM-IF is two orders of magnitude larger than that utilized by the other methods.

Furthermore, we present the confusion matrices for each method, evaluated across all residues, on both the CAMEO dataset (Figure 2) and the CASP15 dataset (Figure A1). The results indicate that OPUS-Design achieves superior performance compared to other methods for the majority of residue types.

To examine the first-stage 3D CNN independently of subsequent modules, we compare its direct residue predictions with those of the 3D CNN-based models ProDCoNN [17] and DenseCPD [18] on the CAMEO dataset (Table A3). Here, all OPUS-Design variants in this comparison use a single 3D CNN pass, without feature aggregation or ESM-2 refinement. OPUS-Design achieves 54.42% recovery for all residues and 65.95% for core residues, compared with 37.91% and 42.59% for ProDCoNN and 43.85% and 52.00% for DenseCPD. Removing side-chain

2) http://github.com/OPUS-MaLab/opus_design.

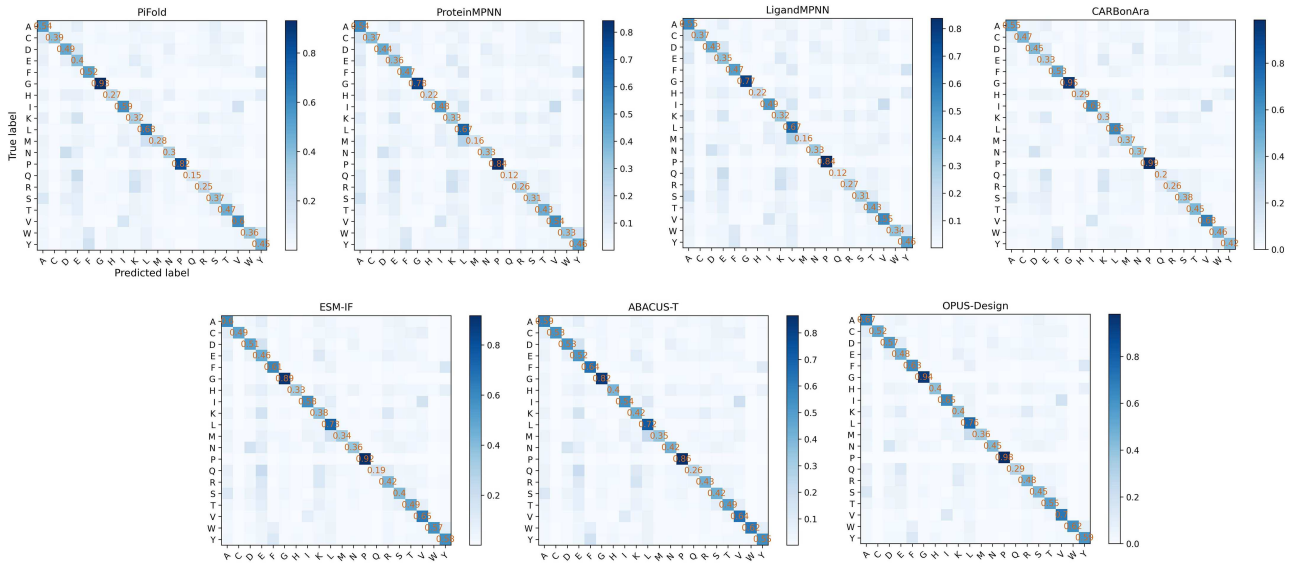

Figure 2 (Color online) The confusion matrices for each method on the CAMEO dataset, evaluated across all residues.

density supervision reduces recovery to 50.83% and 61.81%, respectively. Further removing voxel-occupancy supervision reduces recovery to 49.50% and 59.67%. These results support the effectiveness of the first-stage 3D CNN and show that both side-chain density and voxel-occupancy supervision contribute to its residue prediction.

In practical protein engineering, successful design often benefits from integrating diverse methodologies [43]. Combining predictions from different methods can therefore provide useful guidance alongside the performance of individual methods. By integrating predictions from ESM-IF and OPUS-Design, we find that the two methods predict the same amino acid for 57% of all residues and 67% of core residues in the CAMEO test set. Within these agreement subsets, the sequence recovery rates are 80.37% and 88.37%, respectively. These results highlight the potential value of integrating predictions from different methods to prioritize residue choices for subsequent protein design. However, residues with matching predictions may be enriched for more readily predictable environments, and their recovery rates should therefore be interpreted as conditional on agreement, rather than as representative of all residues or challenging structural microenvironments. Thus, agreement-based selection offers a promising strategy for guiding protein design, although its practical benefits require further validation.

3.2 Performance of different methods on oligomer targets

Since the 3D convolutional network treats each atom separately, OPUS-Design can naturally generalize to multi-chain scenarios even though it is trained exclusively on single chains. We therefore evaluate the performance of different methods on the oligomer test set CAMEO-oligomer. For methods that lack the capability to process multi-chain structures in their code, such as PiFold, we assign the same chain ID to all chains and sequentially adjust the residue IDs to convert the oligomer into a monomer format. As shown in Table 2, OPUS-Design exhibits superior performance relative to the other methods, exhibiting higher sequence recovery rates, both inclusively across all residues and specifically among interfacial residues. In Figure 3, we present the confusion matrices for each method, evaluated across interfacial residues.

3.3 Performance of different methods on ligand-bound complexes

We further evaluate the performance of competing methods on the PXM-2025 dataset of ligand-bound protein complexes. As presented in Table 3, OPUS-Design achieves superior performance over other approaches and yields higher sequence recovery rates for both all residues and ligand-related residues. We also report results for OPUS-Design without ligand input. Its accuracy on ligand-related residues decreases, which indicates the model's sensitivity to local structural environments and its ability to exploit ligand information.

We illustrate four paired examples (Figure A2) to show how ligand input changes predicted residues and modeled pocket contacts, and calculate their Rosetta pocket interaction energies [44]. In Figure A2(a), six ligand-related residues differ between the two predictions. In Figures A2(b) and (c), the predictions made without ligand input show short contacts suggestive of steric clashes. In Figure A2(d), the ligand-aware prediction shows a possible hydrogen bond. Although the ligand-aware designs have lower calculated scores in all four cases, these scores

Table 2 Performance comparison of different methods on the CAMEO-oligomer dataset, measured by all residues and interfacial residues.

Method	Sequence recovery (%) ↑	Charge MAE ↓	Hydropathy MAE ↓	AA JSD ↓	BLOSUM62 ↑	ESM-C PLL ↑
All residues						
PiFold	50.42	0.27	1.43	0.04	2.47	−2.13
ProteinMPNN	46.98	0.28	1.50	0.05	2.30	−1.82
LigandMPNN	51.05	0.26	1.38	0.04	2.54	−1.87
CARBonAra	51.64	0.26	1.24	0.03	2.69	−2.09
ESM-IF	57.22	0.26	1.19	0.05	2.93	−1.71
ABACUS-T	55.01	0.26	1.23	0.05	2.75	−1.66
OPUS-Design	61.61	0.21	1.06	0.03	3.16	−1.73
Interfacial residues						
PiFold	48.20	0.26	1.58	0.16	2.31	−2.37
ProteinMPNN	45.25	0.27	1.61	0.17	2.19	−2.12
LigandMPNN	50.33	0.24	1.48	0.15	2.42	−2.18
CARBonAra	51.26	0.25	1.40	0.14	2.57	−2.32
ESM-IF	55.74	0.23	1.28	0.13	2.87	−2.01
ABACUS-T	51.41	0.24	1.37	0.16	2.60	−1.91
OPUS-Design	60.41	0.20	1.15	0.13	3.07	−1.95

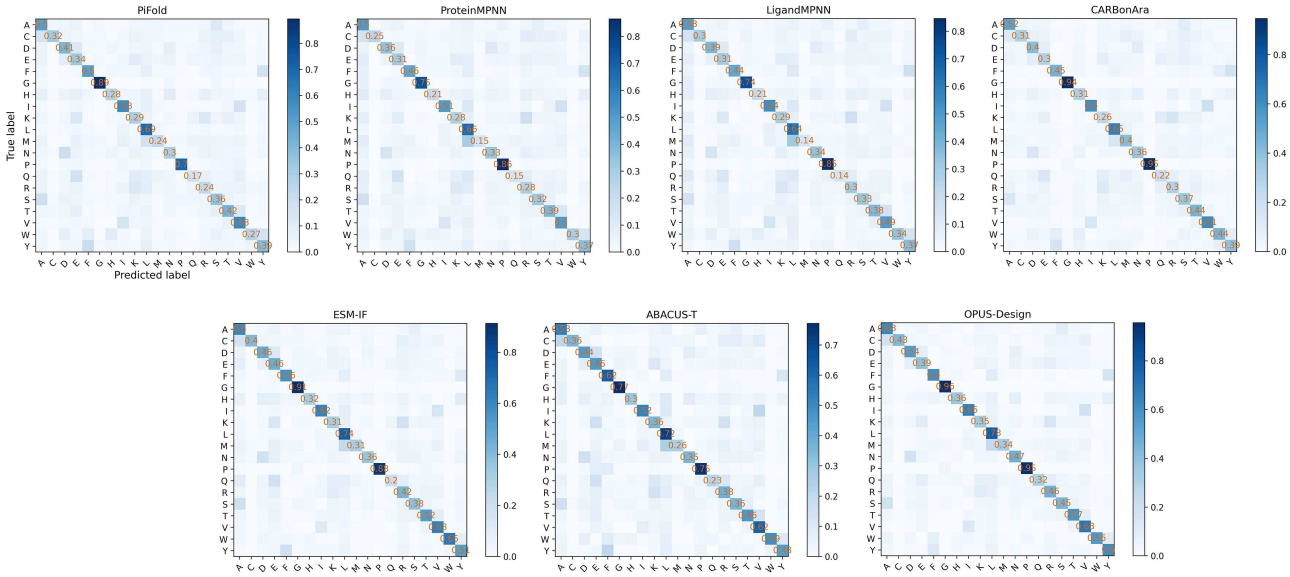

Figure 3 (Color online) The confusion matrices for each method on the CAMEO-oligomer dataset, evaluated across interfacial residues.

depend on modeled side chains and do not measure binding affinity. We therefore interpret the examples as context-dependent local effects, rather than evidence of a general improvement in experimental binding.

Note that OPUS-Design is a general structure-conditioned design model rather than an architecture developed specifically for ligand-bound targets. Its volumetric input groups atoms into five broad classes (C, N, O, S, and others), providing information about spatial occupancy and coarse atom type without explicitly representing ligand identity or detailed bonding chemistry. This limited representation is consistent with the modest aggregate effect of ligand input, while the pocket examples illustrate that ligand atoms can still affect particular predictions. Incorporating more detailed ligand chemistry is a direction for future development.

3.4 Ablation study of OPUS-Design

To assess the contributions of individual components, we conduct ablation studies on all four datasets (Figure 4). Removing the features produced by the 3DCNN feature extraction module causes the largest reduction in sequence recovery among the tested ablations, highlighting the importance of local three-dimensional structural information. Direct prediction from the 3DCNN alone also consistently outperforms prediction using the feature aggregation module alone. Removing the feature aggregation module reduces sequence recovery on all four datasets, whereas

Table 3 Performance comparison of different methods on the PXM-2025 dataset, measured by all residues and ligand-related residues.

Method	Sequence recovery (%) ↑	Charge MAE ↓	Hydropathy MAE ↓	AA JSD ↓	BLOSUM62 ↑	ESM-C PLL ↑
All residues						
PiFold	55.40	0.24	1.22	0.03	2.86	−2.11
ProteinMPNN	48.48	0.26	1.46	0.05	2.45	−1.95
LigandMPNN	55.15	0.23	1.27	0.03	2.86	−1.96
CARBonAra	69.49	0.16	0.75	0.02	3.77	−1.98
ESM-IF	63.48	0.20	1.00	0.04	3.34	−1.75
ABACUS-T	60.44	0.19	1.06	0.03	3.21	−1.69
OPUS-Design	75.24	0.11	0.59	0.01	4.09	−1.68
OPUS-Design (w/o ligand)	75.44	0.11	0.60	0.02	4.07	−1.68
Ligand-related residues						
PiFold	54.07	0.20	1.26	0.09	2.90	−2.13
ProteinMPNN	46.27	0.23	1.53	0.11	2.36	−2.04
LigandMPNN	58.83	0.19	1.17	0.08	3.17	−1.98
CARBonAra	73.77	0.11	0.65	0.05	4.17	−1.86
ESM-IF	65.76	0.15	0.99	0.07	3.56	−1.65
ABACUS-T	64.52	0.12	0.91	0.07	3.60	−1.55
OPUS-Design	79.77	0.07	0.49	0.04	4.41	−1.53
OPUS-Design (w/o ligand)	78.13	0.08	0.55	0.04	4.28	−1.57

removing ESM-2 reduces recovery on CAMEO, CAMEO-oligomer, and PXM-2025 but produces a slight improvement on CASP15. Additional ablations show that removing pairwise contact features, recycling, or IPA, or reducing the number of InvFormer and IPA layers, produces relatively small changes. Together, these results identify 3DCNN-based feature extraction as the principal contributor to the observed performance and highlight that global structural context and evolutionary information may need to be provided by complementary modules to achieve optimal performance.

3.5 Evaluating sequence design methods via structure prediction

In protein sequence design, many studies include a structure prediction comparison task to demonstrate the quality of designed sequences. However, we argue that using structural metrics to evaluate sequence design approaches can be inaccurate, due to inherent limitations of current structure prediction methods [45]. As shown in Figure 5 and Table 4, we use ESMFold [23] to compare the structural similarity between structures predicted from native sequences and those generated from designed sequences. The results indicate that although sequence recovery rates differ considerably across methods, the sc-TM and predicted local distance difference test (pLDDT) scores remain very close. Given the intrinsic noise and errors in structure prediction pipelines, such small differences in structural similarity metrics may not reflect meaningful or reliable distinctions among sequence design methods.

Most importantly, since we have already known the target structure, we can enhance the sc-TM score by iteratively executing the program and selecting the sequence that exhibits the highest structural similarity to the target structure. To incorporate randomness into OPUS-Design, we randomly mask the local environmental features of several residues before introducing them into the feature aggregation module. By iteratively executing the second stage of OPUS-Design 18 times, we obtain 18 distinct sequences. Our findings, as presented in Table 4 (OPUS-Design-rand18), indicate that by selecting the predicted structure with the highest structural similarity to the target structure, OPUS-Design-rand18 achieves the highest sc-TM value of 0.97, albeit with a slight decrease in sequence recovery rate compared to OPUS-Design.

As a complement to the ESMFold analysis, OmegaFold [46] provides an additional structural evaluation (Table A4). OPUS-Design achieves an sc-TM of 0.90, while OPUS-Design-rand18 reaches the highest sc-TM of 0.93 when OmegaFold is used to select the design.

4 Discussion

Protein inverse folding is not only a fundamental task in computational protein design, but also a representative problem in modeling high-dimensional biological complex systems. It centers on deciphering the structural and functional principles encoded within intricate spatial microenvironments, such as those found in protein oligomers

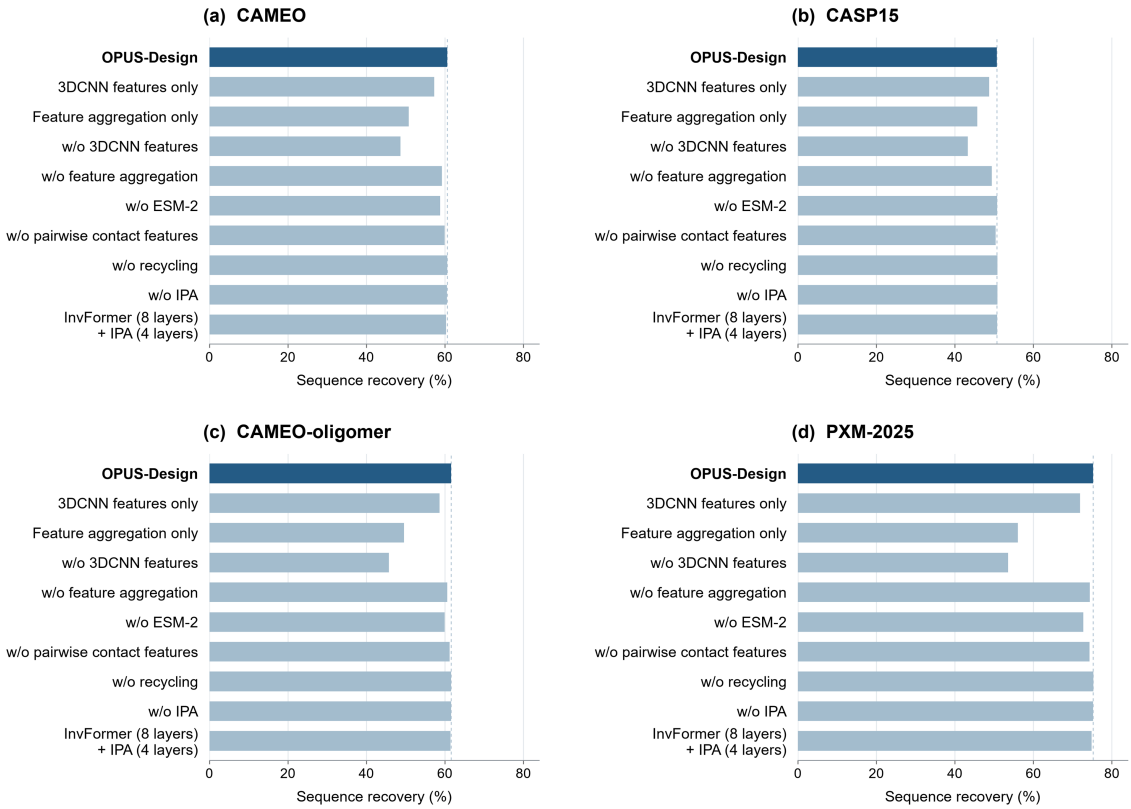

Figure 4 (Color online) Ablation study of OPUS-Design on (a) CAMEO, (b) CASP15, (c) CAMEO-oligomer, and (d) PXM-2025. "3DCNN features" denotes the 3DCNN feature extraction module, and "feature aggregation" denotes the feature aggregation module, which includes InvFormer and IPA. OPUS-Design denotes the complete model. "3DCNN features only" predicts residue identities directly using the 3DCNN prediction head, without feature aggregation or ESM-2. "Feature aggregation only" uses the feature aggregation module without 3DCNN features or ESM-2. "w/o 3DCNN features" excludes the 3DCNN features while retaining feature aggregation and ESM-2. "w/o feature aggregation" removes the entire feature aggregation module, including InvFormer and IPA, while retaining the 3DCNN feature extraction module and ESM-2. The remaining "w/o" variants remove ESM-2, pairwise contact features, recycling within the feature aggregation module, or IPA, respectively. "InvFormer (8 layers) + IPA (4 layers)" reduces the corresponding layer counts to eight and four. Dark blue bars indicate the complete model, and light blue bars indicate ablation variants. Dashed vertical lines mark the recovery of the complete model in each panel. Sequence recovery is calculated as the percentage of correctly predicted residues among all residues. The "3DCNN features only", "w/o feature aggregation", "w/o ESM-2", and "w/o recycling" variants use existing trained weights without additional training, with the corresponding components bypassed or disabled during inference. All other ablation variants are retrained.

Table 4 Comparison of sequence recovery rates, sc-TM scores, and pLDDT values for different methods on the CAMEO-TM90 dataset. Sequence recovery is calculated as the percentage of correctly predicted residues among all residues.

	Sequence recovery (%)	sc-TM	pLDDT
PiFold	53.36	0.93	85.35
ProteinMPNN	48.37	0.95	89.78
LigandMPNN	52.46	0.95	89.54
CARBonAra	54.10	0.95	87.65
ESM-IF	60.67	0.95	88.95
ABACUS-T	59.28	0.96	89.78
OPUS-Design	63.58	0.95	89.13
OPUS-Design-rand18	63.19	0.97	90.12

and ligand-bound complexes. While graph-based architectures currently dominate the inverse folding field and have achieved notable success in modeling discrete protein structural connections, they are not the only viable framework for encoding and reasoning over 3D structural information and complex spatial contexts across diverse protein assemblies. As an alternative, 3DCNNs offer a powerful and complementary paradigm. They excel at capturing dense spatial context and fine-grained details within complex protein spatial microenvironments, possess inherent rotation invariance that is well-suited for structural spatial analysis, and can directly model continuous local environments from volumetric representations. Collectively, these properties ideally satisfy the requirements for multiscale, high-precision characterization of intricate spatial microenvironments in protein systems.

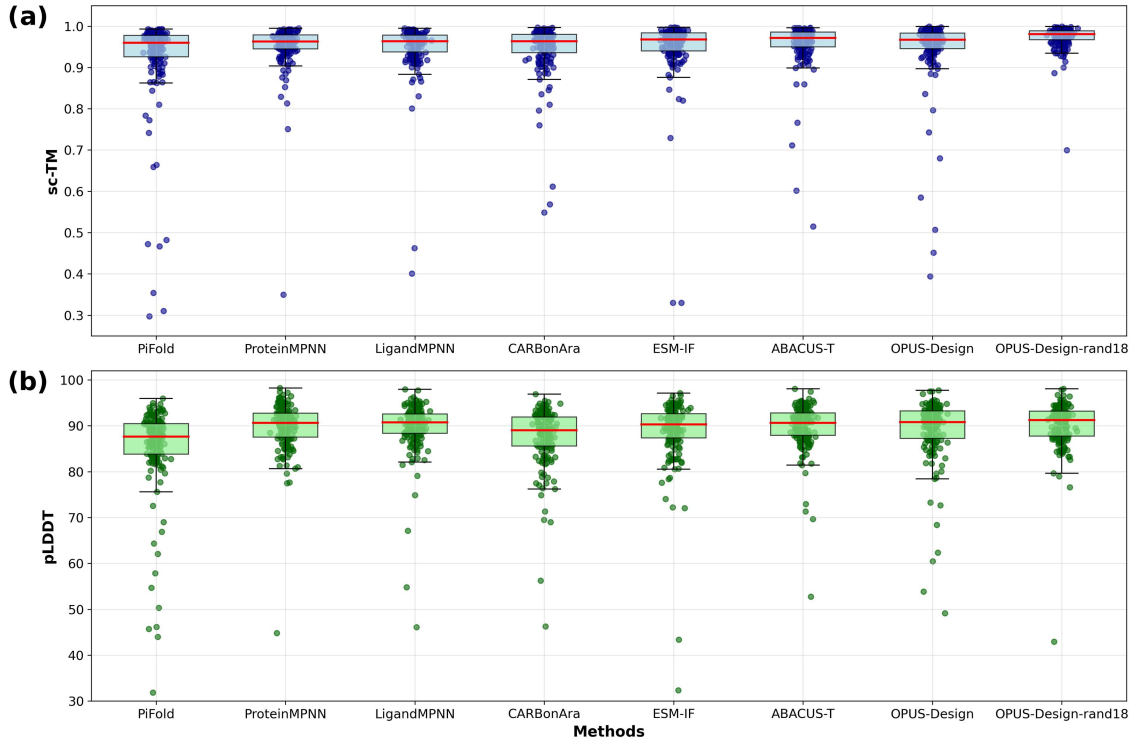

Figure 5 (Color online) The results of structural indicators of each method on the CAMEO-TM90 dataset. (a) The comparison of sc-TMs achieved by different methods; (b) the comparison of pLDDT values achieved by different methods.

To fully unlock the potential of 3DCNNs in addressing the key modeling bottlenecks of complex protein spatial microenvironments, in this paper, we propose OPUS-Design, a two-stage method that discretizes atoms into 3D voxels along with their atomic radii and charges to capture fine-grained local structural environments. It then aggregates these local features together with evolutionary information derived from ESM-2 to predict protein sequences. Our evaluations on monomers (Tables 1 and A2, Figures 2 and A1), oligomers (Table 2 and Figure 3), and protein-ligand complexes (Table 3) using recently released targets demonstrate that OPUS-Design achieves sequence recovery comparable to or even superior to that of other state-of-the-art graph-based methods, thereby validating the efficacy of 3DCNN for high-accuracy protein design.

OPUS-Design also achieves the lowest charge and hydropathy MAE and the highest BLOSUM62 score in most cases. These results support the preservation of native-like residue substitutions and physicochemical properties, including at interfacial and ligand-related sites. Because this study focuses on the contribution of 3DCNN-based modeling to accurate residue-type prediction, the primary OPUS-Design setting produces deterministic predictions rather than sampling multiple sequences per target. Within-target design diversity is therefore not evaluated here and remains a direction for future model development.

Ablation studies (Figure 4) reveal that the 3DCNN module provides a dominant contribution to sequence recovery, while evolutionary information derived from ESM-2 also plays a critical role. Notably, all methods exhibit degraded performance on the CASP15 dataset relative to CAMEO (Table 1), likely due to the low homology of CASP15 targets. In this challenging scenario, OPUS-Design achieves the best overall performance. Moreover, the variant of OPUS-Design without evolutionary information performs even better, underscoring the strong generalization capacity of the 3DCNN backbone.

We further evaluate sequence design methods through structure prediction (Figure 5, Tables 4 and A4). Our results show that while sequence recovery varies noticeably across different methods, the resulting sc-TM and pLDDT scores remain highly comparable. Given the inherent noise and uncertainties present in structure prediction pipelines, we argue that such minor variations in structural similarity metrics may not correspond to meaningful or statistically robust differences between sequence design approaches.

In summary, our findings establish 3DCNN as a competitive alternative to graph-based architectures for protein inverse folding. These results further suggest that volumetric deep learning represents a highly promising and powerful framework for addressing a broad spectrum of protein structure modeling tasks.

Note that, in this paper, we do not claim that our 3DCNN-based method outperforms other method types in

all scenarios. 3DCNN-based, graph-based, and context-aware methods each have their own merits. Our paper showcases that our 3DCNN-based method can achieve results competitive with those of current leading methods, especially in terms of recovery rate; in other words, identifying residues according to their spatial microenvironments, which offers practical insights for model design in the protein design field.

However, 3DCNN also presents several challenges compared with graph-based architectures. For instance, standard 3DCNN lacks the ability to explicitly capture global structural information, thus requiring dedicated aggregation modules to integrate long-range contextual cues for optimal performance. Meanwhile, 3DCNN-based methods often suffer from lower computational efficiency, as each residue must be processed individually, leading to high overhead when updating local environments across all residues. In the future, these limitations need to be addressed to fully unlock the potential of 3DCNN-based frameworks in protein structure modeling.

To assess the practical impact of this overhead, we benchmark OPUS-Design, ProteinMPNN, and ESM-IF on the same 32 GB NVIDIA V100 GPU using representative proteins with 100, 203, and 502 residues. Including input preparation, the estimated OPUS-Design runtimes are 2.7, 5.2, and 13.0 min, respectively, compared with 1.25, 1.44, and 2.60 s for ProteinMPNN and 1.61, 2.40, and 6.05 s for ESM-IF. The corresponding peak GPU memory usage is 15.3, 17.3, and 29.4 GiB for OPUS-Design, 1.0–1.3 GiB for ProteinMPNN, and 2.6–10.0 GiB for ESM-IF. The sequential 3DCNN stage accounts for most of the runtime. The current implementation produces one deterministic sequence per target without repeated sampling. Although batch processing may improve throughput, it delays updates to neighboring local contexts, and our preliminary tests suggest that batch-wise updates reduce sequence recovery. We therefore retain sequential processing to prioritize recovery. Developing more efficient update strategies that preserve accuracy remains a direction for future work.

A further limitation is the absence of experimental validation. Although the additional sequence-based metrics, structure predictions, and representative pocket examples broaden the computational assessment, they cannot guarantee whether the designed proteins fold, bind, or function as intended. Experimental testing of representative designs is therefore an important next step.

Acknowledgements Jianpeng MA wants to thank the support from National Key Research and Development Program of China (Grant No. 2024YFA1307502), Science and Technology Innovation Plan of Shanghai Science and Technology Commission (Grant No. 23JS1400200), Research Fund for International Senior Scientists (Grant No. W2431060), and Research Fund of Development and Reform Commission of Hunan province (Grant No. 2507-430000-04-05-666048). Gang XU wants to thank the support from National Natural Science Foundation of China (Grant No. 32300535).

Supporting information Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Silva D A, Yu S, Ulge U Y, et al. De novo design of potent and selective mimics of IL-2 and IL-15. Nature, 2019, 565: 186–191
2 Correia B E, Bates J T, Loomis R J, et al. Proof of principle for epitope-focused vaccine design. Nature, 2014, 507: 201–206
3 Praetorius F, Leung P J Y, Tessmer M H, et al. Design of stimulus-responsive two-state hinge proteins. Science, 2023, 381: 754–760
4 Vázquez Torres S, Leung P J Y, Venkatesh P, et al. De novo design of high-affinity binders of bioactive helical peptides. Nature, 2024, 626: 435–442
5 Cao L, Coventry B, Goreshnik I, et al. Design of protein-binding proteins from the target structure alone. Nature, 2022, 605: 551–560
6 An L, Said M, Tran L, et al. Binding and sensing diverse small molecules using shape-complementary pseudocycles. Science, 2024, 385: 276–282
7 Bick M J, Greisen P J, Morey K J, et al. Computational design of environmental sensors for the potent opioid fentanyl. eLife, 2017, 6: 28909
8 Huddy T F, Hsia Y, Kibler R D, et al. Blueprinting extendable nanomaterials with standardized protein blocks. Nature, 2024, 627: 898–904
9 Lu H, Diaz D J, Czarnecki N J, et al. Machine learning-aided engineering of hydrolases for PET depolymerization. Nature, 2022, 604: 662–667
10 Siegel J B, Zanghellini A, Lovick H M, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science, 2010, 329: 309–313
11 Reetz M T, Qu G, Sun Z. Engineered enzymes for the synthesis of pharmaceuticals and other high-value products. Nat Synth, 2024, 3: 19–32
12 Wang Q, Liu X, Zhang H, et al. Cytochrome P450 enzyme design by constraining the catalytic pocket in a diffusion model. Research, 2024, 7: 0413
13 Lin Z Y, Li H, Lv L Z H, et al. TaxDiff: taxonomic-guided diffusion model for protein sequence generation. Sci China Inf Sci, 2025, 68: 149101
14 Gao Z, Tan C, Zhang Y, et al. ProteinInvBench: benchmarking protein inverse folding on diverse tasks, models, and metrics. In: Proceedings of Advances in Neural Information Processing Systems 36, 2023. 68207–68220
15 Li Z, Yang Y, Faraggi E, et al. Direct prediction of profiles of sequences compatible with a protein structure by neural networks with fragment-based local and energy-based nonlocal profiles. Proteins, 2014, 82: 2565–2573
16 O'Connell J, Li Z, Hanson J, et al. SPIN2: predicting sequence profiles from protein structures using deep neural networks. Proteins, 2018, 86: 629–633
17 Zhang Y, Chen Y, Wang C, et al. ProDCoNN: protein design using a convolutional neural network. Proteins, 2020, 88: 819–829
18 Qi Y, Zhang J Z H. DenseCPD: improving the accuracy of neural-network-based computational protein sequence design with densenet. J Chem Inf Model, 2020, 60: 1245–1252
19 Gao Z, Tan C, Li S Z. PiFold: toward effective and efficient protein inverse folding. In: Proceedings of the 11th International Conference on Learning Representations, 2023

20 Dauparas J, Anishchenko I, Bennett N, et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science, 2022, 378: 49–56
21 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, 2022. 8946–8970
22 Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596: 583–589
23 Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science, 2023, 379: 1123–1130
24 Ren M, Yu C, Bu D, et al. Accurate and robust protein sequence design with CarbonDesign. Nat Mach Intell, 2024, 6: 536–547
25 Liu Y, Wu R, Wang X, et al. Enhancing functional proteins through multimodal inverse folding with ABACUS-T. Nat Commun, 2025, 16: 10177
26 Qin Z R, Lin M B, Liu H B, et al. Low-rank Winograd transformation for 3D convolutional neural networks. Sci China Inf Sci, 2025, 68: 159101
27 Cheng D Z, Zhang X. Semi-tensor product-based convolutional neural network. Sci China Inf Sci, 2026, 69: 140201
28 Ye Z M, Zheng H, Han G, et al. CFAN: convolutional frequency-attention network for ECG-based emotion recognition. Sci China Inf Sci, 2026, 69: 162204
29 Xu G, Luo Z, Yan Y, et al. OPUS-Rota5: a highly accurate protein side-chain modeling method with 3D-Unet and RotaFormer. Structure, 2024, 32: 1001–1010
30 Misiura M, Shroff R, Thyer R, et al. DLPacker: deep learning for prediction of amino acid side chain conformations in proteins. Proteins, 2022, 90: 1278–1290
31 Wang J, Cieplak P, Kollman P A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem, 2000, 21: 1049–1074
32 Kingma D P, Ba J. Adam: a method for stochastic optimization. ArXiv:1412.6980
33 Yang J, Anishchenko I, Park H, et al. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci USA, 2020, 117: 1496–1503
34 Haas J, Barbato A, Behringer D, et al. Continuous automated model evaluation (CAMEO) complementing the critical assessment of structure prediction in CASP12. Proteins, 2018, 86: 387–398
35 Protenix Team, Zhang Y, Gong C, et al. Protenix-v1: toward high-accuracy open-source biomolecular structure prediction. bioRxiv, 2026. https://www.biorxiv.org/content/10.64898/2026.02.05.703733v1.full.pdf
36 Lin J. Divergence measures based on the Shannon entropy. IEEE Trans Inform Theor, 1991, 37: 145–151
37 Henikoff S, Henikoff J G. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA, 1992, 89: 10915–10919
38 Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein. J Mol Biol, 1982, 157: 105–132
39 Candido S, Hayes T, Derry A, et al. Language modeling materializes a world model of protein biology. bioRxiv, 2026. https://www.biorxiv.org/content/10.64898/2026.06.03.729735v1
40 Huang X, Pearce R, Zhang Y, et al. FASPR: an open-source tool for fast and accurate protein side-chain packing. Bioinformatics, 2020, 36: 3758–3765
41 Dauparas J, Lee G R, Pecoraro R, et al. Atomic context-conditioned protein sequence design using LigandMPNN. Nat Methods, 2025, 22: 717–723
42 Krapp L F, Meireles F A, Abriata L A, et al. Context-aware geometric deep learning for protein sequence design. Nat Commun, 2024, 15: 6273
43 Johnson S R, Fu X, Viknander S, et al. Computational scoring and experimental evaluation of enzymes generated by neural networks. Nat Biotechnol, 2025, 43: 396–405
44 Alford R F, Leaver-Fay A, Jeliazkov J R, et al. The Rosetta all-atom energy function for macromolecular modeling and design. J Chem Theor Comput, 2017, 13: 3031–3048
45 Garcia M, Dixit S M, Rocklin G J. Evaluating zero-shot prediction of monomeric protein design success by AlphaFold, ESMFold, and ProteinMPNN. Protein Sci, 2026, 35: 70453
46 Wu R, Ding F, Wang R, et al. High-resolution de novo structure prediction from primary sequence. bioRxiv, 2022. https://www.biorxiv.org/content/10.1101/2022.07.21.500999v1