CryoDomain: Sequence-free Protein Domain Identification from Low-resolution Cryo-EM Density Maps

Muzhi Dai, Zhuoer Dong, Weining Fu, Kui Xu *, Qiangfeng Cliff Zhang *

School of Life Sciences, Tsinghua University, Beijing, 100084, China
{dmz21,dze18,fwn23}@mails.tsinghua.edu.cn, {xukui,qczhang}@tsinghua.edu.cn

Abstract

Cryo-electron microscopy (cryo-EM) has revolutionized the field of structural biology, determining structures of large protein machines and sharpening the understanding of fundamental biological processes. Despite cryo-EM's unique capacity to discover novel proteins from unpurified samples and reveal the intricate structures of protein complexes within native cellular environments, the advancement of protein identification methods for cryo-EM lags behind. Without prior knowledge, such as sequence, protein identification from low-resolution density maps remains challenging. Here we introduce CryoDomain, an innovative method for identifying protein domains — conserved constituent units of proteins — from low-resolution cryo-EM density maps without requiring prior knowledge of protein sequences. CryoDomain leverages cross-modal alignment to correlate cryo-EM density maps with atomic structures, transferring the knowledge learned on a large atomic structure dataset to a sparse density map dataset. On two protein domain benchmarks constructed from CATH and SCOPe, CryoDomain significantly outperforms the state-of-the-art methods for domain identification from low-resolution density maps. CryoDomain liberates structural biologists from the tedious tasks of density inspection and database searching during protein identification. It has the potential to extend the border of unbiased structure discovery and cellular landscape investigation using cryo-EM.

Introduction

Structure determination of proteins is crucial to understanding biological processes and molecular mechanisms (Liao et al. 2013)(Taylor et al. 2015). Cryo-EM has emerged as a popular approach for structure determination of endogenous macromolecules (Kühlbrandt 2014)(Nogales and Mahamid 2024), with the number of cryo-EM density maps deposited in the Electron Microscopy Data Bank (EMDB) growing exponentially (Lawson et al. 2016). Furthermore, a burgeoning subdivision of cryo-EM, cryo-electron tomography (cryo-ET), is capable of revealing *in situ* nanometer-scale structural information of cellular components in their native environments, while indispensable molecular interactions might be lost in traditional single particle analysis (SPA) cryo-EM (Turk and Baumeister 2020).

*Corresponding authors.

Atomic-resolution protein structure determination from cryo-EM demands high-resolution density maps and usually requires primary amino acid sequences. As the resolution of density map drops, density regions corresponding to protein side chains become blurred and hardly recognizable (Fraser and Murcko 2024). In extreme cases, a density region encompassing several amino acids might be completely lost, making it difficult or even impossible to identify the chemical identity of individual amino acids. Without prior knowledge such as type or sequence of the protein, structure determination methods are generally not applicable to low-resolution density maps. Thus, protein identification in a broader perspective turns out to be an important alternative for the interpretation of density maps. Sequence search-based methods (Ho et al. 2020)(Jamali et al. 2024) first recognize individual amino acids from a density map, and then build a stretch of amino acids for subsequent sequence searching. Structure search-based approaches (Chang et al. 2022b) can also achieve protein identification, provided that a sufficient fraction of backbone structure can be built from the density map. However, both approaches still suffer from sharp performance decline on low-resolution (>4 Å) density maps since they heavily rely on accurate amino acid recognition. Meanwhile, to improve accuracy and efficiency for short sequence or cracked backbone identification, these approaches often restrict the search space by resorting to other biological experiments such as tandem mass spectrometry (MS/MS) which can detect all proteins in the sample of interest (Chang et al. 2022a)(Gui et al. 2022). The recent revolution in sequence-based structure prediction (Jumper et al. 2021) leads to another option, i.e., predicting the structure for each protein in the search space and sorting by the correlation with the density map, but this is quite time-consuming and limited by the quality of predicted structures (Gao et al. 2023).

Different from amino acid-dependent methods mentioned above, we innovatively propose a cross-modal method, CryoDomain, for protein identification based on protein domain, especially from low-resolution density maps. Domain is a large constituent unit of proteins, containing 50-250 amino acids (Buljan and Bateman 2009). Missing density or blurred shapes in low-resolution maps has less influence in proportion on domain density regions compared with individual amino acids. In addition, domains exhibit well-conserved structures and functions, where new proteins are merely novel combina-

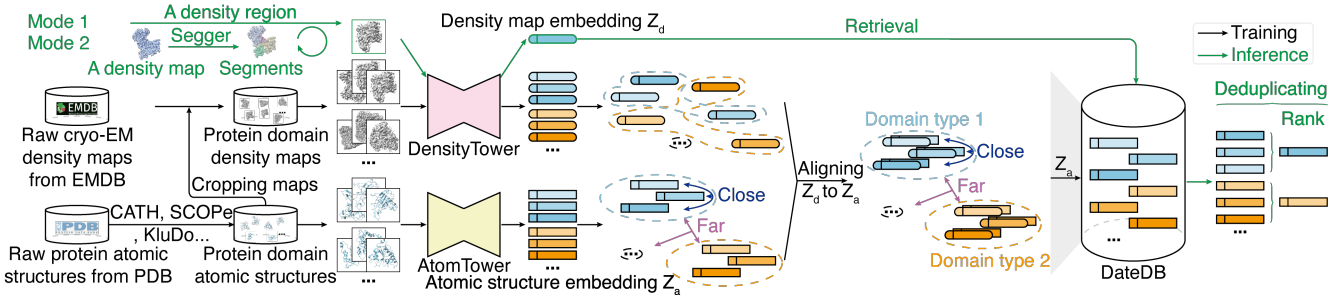

Figure 1: CryoDomain framework. CryoDomain comprises the DensityTower and AtomTower, to simultaneously learn embeddings from protein domain density maps and atomic structures in a shared low-dimension space. The training procedure aligns density map embeddings (Z_d) and atomic structure embeddings (Z_a) of the same domain types. After training, all atomic structure embeddings (Z_a) are stored in the DateDB. Training data are all from publicly available databases, EMDB and PDB (Protein Data Bank) (Berman et al. 2000), and their domain information is from CATH/SCOPe database and KluDo's (Taheri-Ledari et al. 2022) predictions (details in ***Experimental Setup***). During inference, there are two modes in CryoDomain which are guided by green arrows. Mode 1 allows a user-specific density region as input. Mode 2 allows a raw density map as input. The Mode 2 of CryoDomain integrates the density map segmentation tool – Segger (Pintilie et al. 2010) (details in ***Related Works*** and ***Appendix H***). CryoDomain then processes the density map through the DensityTower to generate Z_d, and then queries the DateDB to find the nearest Z_a, ultimately achieving protein domain identification.

tions of existing domains, making dividing and conquering the problem of protein identification into protein domain identification reasonable. To address the disparity in data richness between density maps and atomic structures, and to maximize data utilization, CryoDomain employs a cross-modal framework to integrate these two modalities. Our cross-modal alignment approach effectively distills knowledge learned from rich atomic structure-modal to sparse density map-modal. Cross-modal alignment is achieved by maximizing the similarity between embeddings of the same domain types and minimizing the similarity between embeddings of different domain types. After network training of CryoDomain, we compile an embedding database of known protein domain atomic structures named Density-atom embedding Database (DateDB). Domain identification can be achieved by querying this database with density map embeddings generated by CryoDomain. On two benchmarks constructed from the CATH (Greene et al. 2007) and SCOPe (Fox, Brenner, and Chandonia 2014) databases, CryoDomain significantly outperforms previous methods, such as cryoID (Ho et al. 2020) and ModelAngelo (Jamali et al. 2024), on identifying protein domains from low-resolution density maps. Overall, the main contributions of this work are described as follows:

- We propose a sequence-free protein domain identification method applicable to low-resolution cryo-EM density maps, whereas previous works mainly succeed on high-resolution density maps, and needs additional information such as primary amino acid sequences.
- We adopt the cross-modal alignment strategy to align embeddings of density maps to that of atomic structures, which solves the problem of disparity in data richness between different modalities and enables domain identification via embedding retrieval.
- Extensive experiments conducted on datasets built from CATH and SCOPe showcase that CryoDomain achieves a Top-1 increase of 84% and a Mean Average Precision (mAP) improvement of 67% at least over the existing methods for domain identification at low resolution.

Related Works

There are mainly two related works:

Protein Identification from Density Maps. Preceding methods for protein identification from density maps can be categorized into sequence search-based approaches and structure search-based approaches. Sequence search-based approaches rely on high-resolution density maps that clearly display continuous backbone and accurate side chains. The recognized backbone atoms can be traced into sequence segments, which can be used to search against the sequences of all possible proteins, as done in cryoID. In addition, ModelAngelo adopts hidden Markov model (HMM) which predicts a probability vector for each amino acid and converts these vectors into an HMM profile suitable for searching databases. Instead of local high-quality segments, structure search approaches utilize the entire backbone. DeepTracer-ID (Chang et al. 2022b) predicts a backbone structure from a density map by recognizing and tracing backbone atoms, and then searches AlphaFold2 (Jumper et al. 2021) predicted structures of all possible proteins for candidates. Here, we compared CryoDomain with two open-source methods: ModelAngelo and cryoID, with ModelAngelo being the state-of-the-art.

Density map segmentation. Segmenting raw density maps is a standard and straightforward procedure. Structural biologists always manually segment density regions of interest. There are also several automatic tools to segment a raw density map into multiple regions with little prior knowledge. For instance, Segger (Pintilie et al. 2010) divides density maps based on local maximum regions and iteratively groups them to obtain desired segments. However, annotations (i.e. domain types) are not provided for these segments by Segger. Therefore, identifying the components of the segments remains a challenging task, which CryoDomain focuses on.

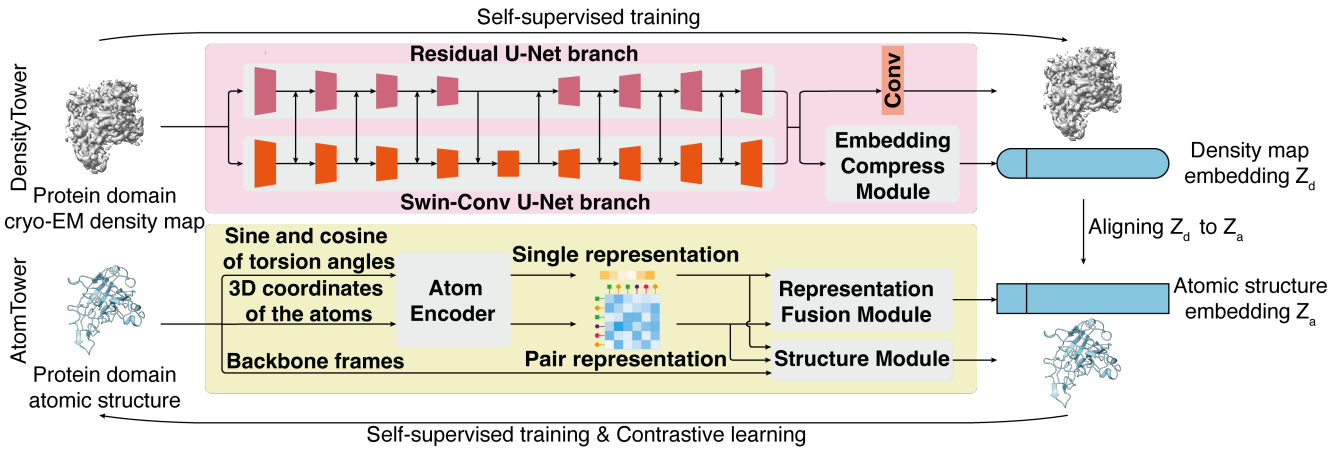

Figure 2: CryoDomain dual-tower network. The DensityTower, comprising a Residual U-Net branch, a Swin-Conv U-Net branch, a *Conv* module, and a Compress Module, extracts a density map embedding (Z_d) from an input density map. The AtomTower, comprising an AtomEncoder, a Structure Module, and a Fusion Module, extracts an atomic structure embedding (Z_a) from an input atomic structure. Both the DensityTower and the AtomTower are self-supervised and pre-trained. Additionally, the AtomTower continues to be trained using contrastive learning. **Torsion angles:** the rotation of one atom group respect to another atom group inside an amino acid (details in ***Appendix D***). **Backbone frames:** the backbone frames record Euclidean transformations from a local coordinate system defined by each amino acid to the global protein coordinate system. Specifically, the local coordinate system is centered on C^{α} atom, with N and C atoms defining X and Y axis (details in ***Appendix D***).

Method

We introduce CryoDomain, a method leveraging a dual-tower deep neural network for identifying protein domains from low-resolution cryo-EM density maps (Figure 1). CryoDomain integrates density maps and atomic structures through cross-modal alignment in a low-dimension space. After aligning the two modalities, CryoDomain constructs a Density-atom embedding Database (DateDB) containing atomic structure embeddings, and CryoDomain identifies protein domains from density maps by embedding retrieval.

CryoDomain Dual-tower Network

We have designed a dual-tower (DensityTower and AtomTower) network architecture for CryoDomain, and realized the alignment of cross-modal representations. As shown in Figure 1 and 2, in the training phase, the DensityTower and AtomTower undergo self-supervised pre-training on raw cryo-EM density maps and atomic structures separately to extract their modality-specific features. The network architecture subsequently integrates the two modalities within a unified representation space through alignment. Training data are all from publicly available databases, EMDB and PDB (Protein Data Bank) (Berman et al. 2000), and their domain information is from CATH/SCOPe database and KluDo's (Taheri-Ledari et al. 2022) predictions (details in ***Experimental Setup***). In the inference phase, the DensityTower receives a density map as input to generate the query embedding. The query embedding is then searched against the pre-generated DateDB to accomplish identification of protein domains.

The DensityTower employs a U-Net-like architecture that progressively learns the spatial and semantic features from cryo-EM density maps. The network consists of two parallel branches, namely a Residual U-Net (He et al. 2016)(Ronneberger, Fischer, and Brox 2015) branch and a Swin-Conv U-Net branch (Zhang et al. 2023), a *Conv* module, and a Compress Module (details in ***Appendix B***). The density map is processed through the Residual and Swin-Conv U-Net branches to incorporate their local and non-local semantic patterns. It is then reconstructed by the *Conv* module and projected into an embedding (Z_d) through the Compress Module. Z_d is then aligned with its corresponding atomic structure embedding (Z_a) during the training phase, and is also used to query pre-generated DateDB for protein domain identification during the inference phase.

The AtomTower, inspired by AlphaFold2 (Jumper et al. 2021), embeds an input atomic structure into single and pair representations through AtomEncoder, and interfuses them with the Fusion Module to form an atomic structure embedding (Z_a) as final output (details in ***Appendix B***). The Structure Module is adapted for recovery of input atomic structures given the single and pair representations (self-supervised training). Additionally, contrastive learning is implemented to enhance the uniqueness of embeddings of different domain types. Two different towers are aligned using mean squared error (MSE) loss, which means that output embeddings (Z_d and Z_a) are unified in a shared space. Consequently, Z_d can be used to retrieve atomic structure embeddings from DateDB for protein identification.

Modality Alignment

Cross-modal alignment enables the integration of heterogeneous data types, such as cryo-EM density maps and atomic structures, crucial for the accuracy of protein domain identification. Due to the sparse pair data of two modalities, traditional cross-modal alignment methods which rely on pair

data for training, such as CLIP (Radford et al. 2021), cannot fully utilize those unpaired portion of data from each modality. Therefore, CryoDomain decomposes the alignment process into domain semantic learning and direct alignment of the two modalities, This training process has three steps:

Self-supervised learning. Firstly, each tower of CryoDomain is trained in a self-supervised manner. This phase allows the network to learn the intricate details of each input modality independently. For the DensityTower, we employ MSE loss for training the network, while the AtomTower utilizes a combination of FAPE (Frame Aligned Point Error), torsion angle, distogram, and pLDDT (predicted local distance difference test) loss (Equation 1∼5 in ***Appendix F***).

Contrastive learning. Secondly, to embed protein domain information explicitly into the atomic structure embeddings (Z_a), AtomTower undergoes contrastive learning. This process ensures that atomic structure embeddings of the same protein domains are closer, while otherwise further, which is the foundation of alignment in a shared domain-specific low-dimension space. The rich annotations of protein domain types for atomic structures underpins the contrastive learning of AtomTower, unlike the sparsely labeled density maps. Here we employ a classic contrastive loss function (Equation 6 in ***Appendix F***) which increases the discrimination of embeddings between different domains by comparing the MSE among positive and negative samples.

Cross-modal alignment. Finally, we use MSE as the alignment loss to map the density map embeddings (Z_d) directly onto the domain-informative atomic structure embeddings (Z_a), so that the two modalities are aligned in a shared domain-specific low-dimension space.

Construction of Density-Atom Embedding Database

In construction of the Density-Atom Embedding Database (DateDB), we first apply the pre-trained AtomTower to extract atomic structure embeddings (Z_a) for structures of all domains (Figure 3). These embeddings are then stored in the DateDB along with pointers to their corresponding atomic structures and other metadata. This repository capacitates subsequent retrieval processes, enabling efficient matching of protein domains based on their atomic structure features.

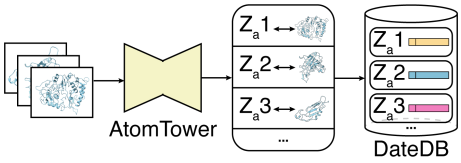

Figure 3: The construction of DateDB. The AtomTower is employed to extract atomic structure embeddings from a protein domain atomic structure dataset. These embeddings, along with their corresponding atomic structures, are systematically stored in the DateDB.

Protein Domain Identification through Retrieval

After cross-modal alignment in a shared domain-specific low-dimension space, we utilize the density map embeddings (Z_d) obtained by DensityTower to retrieve the corresponding atomic structure embeddings. By evaluating the similarity between the Z_d and each Z_a stored in DateDB, we retrieve the top-k atomic structure embeddings most similar to Z_d. The retrieval algorithm employed is the Approximate Nearest Neighbor algorithm and we implemented it by Faiss (Douze et al. 2024). Finally, using the relational data in DateDB, we attain protein domain information of these Z_a, deduplicate the results by only reserving a representative having the maximum cosine similarity with the query for each domain type, and then calculate the final rank (Figure 4).

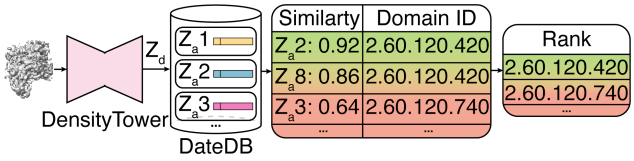

Figure 4: CryoDomain retrieval framework. We employ the DensityTower to extract the density map embedding (Z_d), and use it to retrieve the closest atomic structure embeddings (Z_a) from DateDB. Finally, retrieval results are deduplicated and ranked according to similarity.

Experiments

Experimental Setup

Dataset construction. 1) We downloaded raw cryo-EM density maps with resolutions of 1∼20 Å from EMDB and the corresponding protein atomic structures from PDB, as of April 2023. We also downloaded protein structures solved by other determination methods such as X-ray crystallography. **2)** Since a protein may contain multiple domains, we need to segment atomic structures into individual domains. CATH and SCOPe are two databases (with partial overlap) that specify types and locations of domains (Figure 5 and ***Appendix A***). Metadata in CATH/SCOPe are used to segment atomic structures into protein domains. This step produced about 400k domain structures labelled by CATH and 250k by SCOPe. For those atomic structures from PDB without CATH/SCOPe labels (approximately 50%), we used KluDo to segment them into protein domains. Since KluDo only predicts the positions of domains and does not provide domain types, the domain atomic structures segmented by KluDo are treated as unlabeled ones. **3)** We cropped raw density maps into density regions using the above protein domain structures and removed redundancy, obtaining about 200k final domain density maps (***Appendix C***).

To evaluate CryoDomain, we used pairs of atomic structures and density maps with CATH/SCOPe labels obtained above, and reserved the domain types that contain at least ten pairs to ensure statistical significance. There are 8,780 structures with 159 CATH domain types and 3,038 structures of 79 SCOPe domain types and we used them to build DateDB. Two test sets are built with those low-resolution maps (>4 Å), containing 159 density maps with CATH labels and 129 with SCOPe labels respectively. We guaranteed that domains

in the test sets share no more than 30% sequence identity with those used for training.

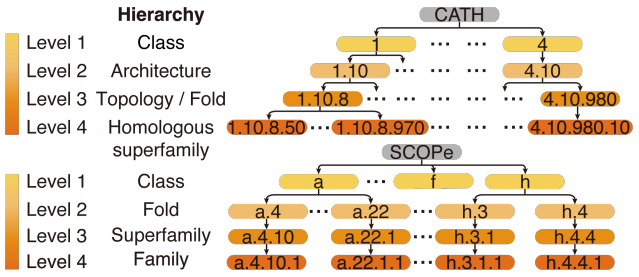

Figure 5: The classification hierarchy of protein domain in CATH/SCOPe. CATH/SCOPe independently classifies protein domains into four hierarchy levels, which indicate the structural and functional similarity between protein domains. Higher hierarchy levels suggest closer relationships.

Baseline methods. We compared CryoDomain with cryoID and ModelAngelo (*without sequence mode*), on protein domain identification from low-resolution density maps. cryoID and ModelAngelo are two methods to identify proteins from density maps and ModelAngelo is the state-of-the-art method. We collected sequences of protein domains with embeddings in CATH/SCOPe DateDB, converted them into sequence databases supplied for cryoID and ModelAngelo, and designated the domain types of returned sequences as identification results (details in ***Appendix G***).

Alignment for density maps and atomic structures

Figure 6 showed the distribution of embeddings of four CATH domain types after each training step. After self-training of the AtomTower and DensityTower, clusters of different domain types can hardly be distinguished (Panel A and B). In comparison, contrastive learning helped CryoDomain separate atomic structure embeddings of different domain types (Panel A and C). After the final cross-modal alignment, embeddings of density maps tended to converge towards those of the corresponding atomic structures (Panel D and C). And density map embeddings of different domains no longer intermingled (Panel D and B), indicating that our cross-modal alignment **1)** unified embeddings of two modalities within the same low-dimensional space, and **2)** ensured that embeddings of different domain types were distinct, while those of the same domain types were similar.

Significant decrease in MSE between two modalities' embeddings on the whole test set (from 3.56 to 0.24) demonstrated that the two modalities are closer after alignment. Concurrently, Top-1/5/10 accuracy and mAP dropped when CryoDomain was trained without alignment, substantiating effectiveness of the alignment (details in ***Ablation study***).

Protein Domain Identification from Low-resolution Density Maps

We evaluated the performance of CryoDomain on protein domain identification using ROC curves, where true positive (TP) and false positive (FP) are defined as atomic structures with correct and wrong domain types (hierarchy level 4) in the

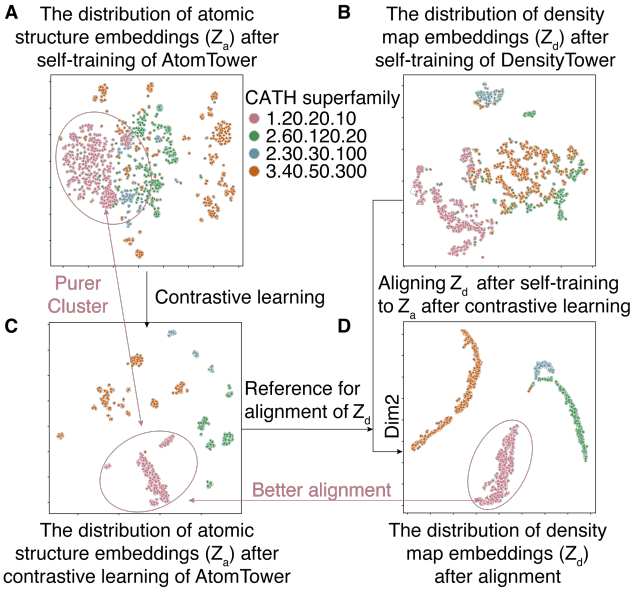

Figure 6: Visualization of the atomic structure and density map embeddings of four CATH domain types using t-SNE (Van der Maaten and Hinton 2008) after each training step.

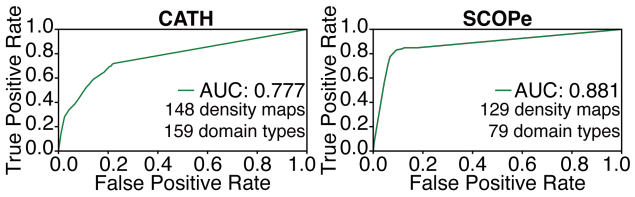

Figure 7: ROC for protein domain identification from low-resolution density maps on the CATH/SCOPe test sets.

list returned by CryoDomain at a given cutoff. CryoDomain achieved an AUROC of over 0.75 for both the CATH and SCOPe test sets, demonstrating its capability of identifying domains from low-resolution maps (Figure 7).

Comparison with other methods. We computed the Top-K Accuracy and Mean Average Precision (mAP) (Equation 7∼9 in ***Appendix F***). As shown in Figure 8, CryoDomain surpassed cryoID on all metrics and performed better than ModelAngelo on all metrics except for the Top-1 Accuracy on the CATH test set which was compensated by higher Top-5 Accuracy. And the mAP of CryoDomain was also higher than that of cryoID and ModelAngelo (Table 1, 0.15 v.s. 0.01 and 0.09 for CATH and 0.47 v.s. 0.01 and 0.01 for SCOPe). The accuracy of Top-1, 5, and 10 of CryoDomain increased progressively, while those of cryoID and ModelAngelo were almost unchanged. Thus, CryoDomain is much more accurate and robust than both cryoID and ModelAngelo and can hit the target protein domain in the Top-5/10 even if it misses in the Top-1, which is acceptable for structural biologists. The CATH/SCOPe hierarchy reflects the evolutionary relationships of protein domains. Although the identification at hierarchy level 4 misses, hitting at lower hierarchy levels can

Method	Dataset	K=1 (%)		K=5 (%)		K=10 (%)		mAP
		4∼5 Å	5∼10 Å	4∼5 Å	5∼10 Å	4∼5 Å	5∼10 Å	4∼10 Å
cryoID	CATH	2.25 (2/89)	0.00 (0/59)	7.87 (7/89)	0.00 (0/59)	7.87 (7/89)	0.00 (0/59)	0.01
ModelAngelo	CATH	**49.44 (44/89)**	0.00 (0/59)	50.56 (45/89)	0.00 (0/59)	50.56 (45/89)	0.00 (0/59)	0.09
CryoDomain	CATH	7.87 (7/89)	**5.08 (3/59)**	**56.18 (50/89)**	**18.64 (11/59)**	**76.40 (68/89)**	**35.59 (21/59)**	**0.15**
cryoID	SCOPe	0.00 (0/2)	11.02 (14/127)	0.00 (0/2)	25.98 (33/127)	0.00 (0/2)	28.35 (36/127)	0.01
ModelAngelo	SCOPe	**50.00 (1/2)**	0.00 (0/127)	**50.00 (1/2)**	0.00 (0/127)	**50.00 (1/2)**	0.00 (0/127)	0.01
CryoDomain	SCOPe	0.00 (0/2)	**57.48 (73/127)**	**50.00 (1/2)**	**74.02 (94/127)**	**50.00 (1/2)**	**88.98 (113/127)**	**0.47**

Table 1: Top-K Accuracy (↑) and mAP (↑) of different methods on the CATH/SCOPe test sets.

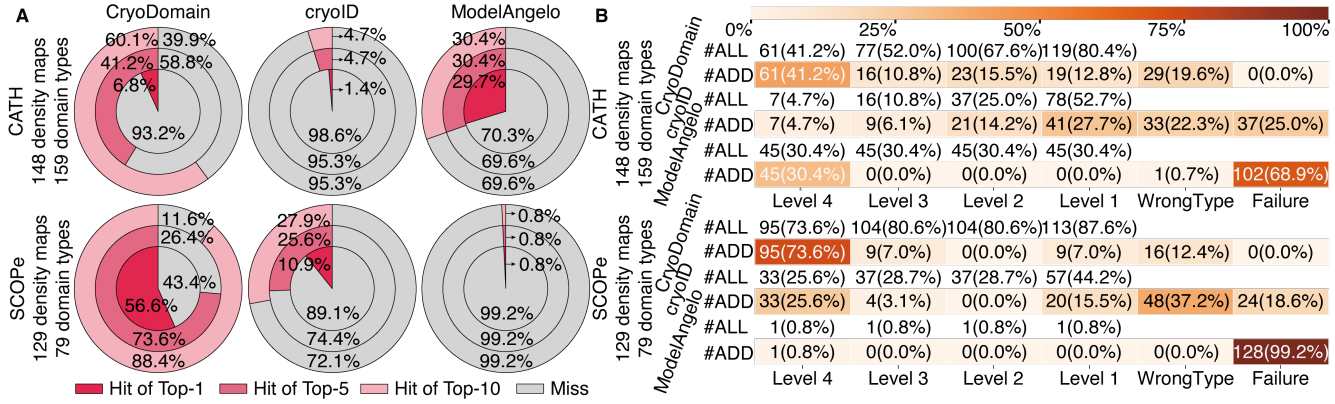

Figure 8: **A:** Top-1, 5, and 10 Accuracy and Miss Rate for protein domain identification at hierarchy level 4 on the CATH/SCOPe test sets. **B:** Top-5 Accuracy at four hierarchy levels on the CATH/SCOPe test sets. For #ALL rows, *Level 1∼4* columns represent that the target domain is hit in the Top-5. For #ADD rows, *Level 3* means that the target protein domain is hit in the Top-5 at hierarchy level 3 while missed at hierarchy level 4; the similar for *Level 4*, *2*, and *1*; *WrongType* means the target protein domain is missed in the Top-5 even at hierarchy level 1; *Failure* refers to null results.

still provide useful insights for biologists. Thus, we also computed the accuracy at hierarchy level 3, 2, and 1 for Top-5 (Figure 8), and found CryoDomian outperformed cryoID and ModelAngelo, further illustrating CryoDomian's accuracy.

The impact of resolution. To explore the impact of density map resolution, we split the test sets into 4∼5 Å ones and 5∼10 Å ones (Table 1). We found that cryoID could not accurately identify protein domains in most cases at both 4∼5 Å and 5∼10 Å resolution ranges. And although ModelAngelo identified protein domains for half of density maps at 4∼5 Å resolution, we found a sharp drop in performance for ModelAngelo in constrast with the slight decline for CryoDomain. To further explore the relationship between resolution and identification accuracy, we low-pass filtered 45 successful cases of ModelAngelo in the Top-5 from the CATH test set to generate low-resolution density maps. Specifically, we ran the cryoSPARC *low-pass filter* tool (Punjani et al. 2017) on 34 of 45 density maps with the resolution parameter of 5, 6, 7, 8, 9, and 10 Å, after removing 11 density maps with execution errors. CryoDomain achieved consistent identification performance for maps with 4∼8 Å resolutions, but showed a gradual decline at 9 Å and 10 Å (Figure 9). In contrast, cryoID got stably poor performance and ModelAngelo suffered a dramatical performance decline at 5 Å with merely 11

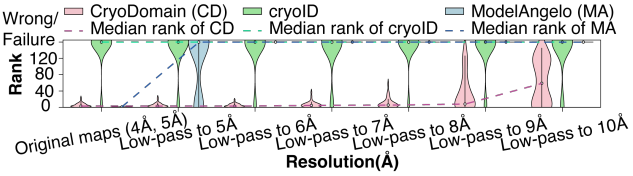

Figure 9: Resolution impact on low-pass filtered maps.

out of 34 success cases in the Top-5 and completely failed at lower resolutions. These indicate that CryoDomain is more accurate and robust over a much broader range of resolutions, significantly outperforming cryoID and ModelAngelo.

Results explanation. The aforementioned results demonstrate that CryoDomain is more accurate and robust on low-resolution density maps compared to cryoID and ModelAngelo. This is consistent with the design principle that recognizing domains is better at resisting the impact of variable noise in low-resolution density maps than those amino acid-dependent methods. However, for density maps with higher resolution (4∼5 Å), CryoDomain got lower Top-1 Accuracy than ModelAngelo, even though its Top-5 Accuracy was higher, indicating that CryoDomain may have a higher false positive rate. ModelAngelo's low false positive rate benefits

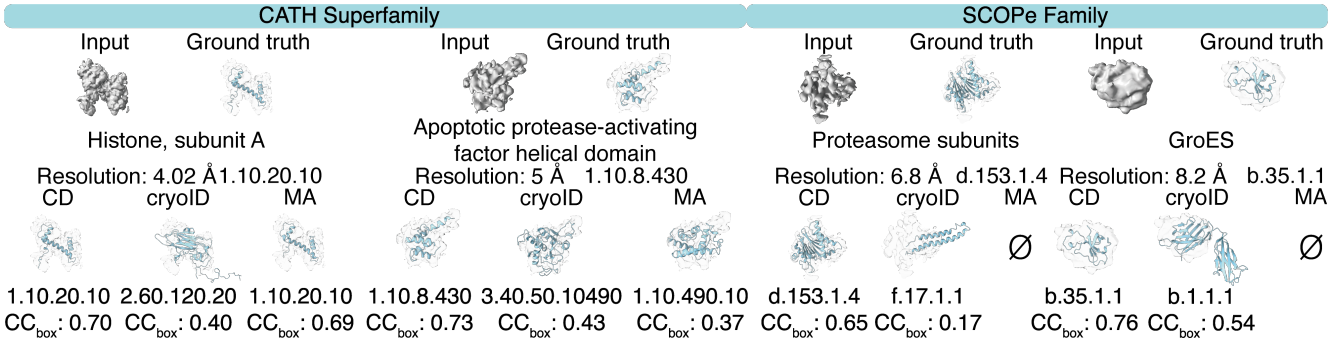

Figure 10: Four cases of domain identification of CryoDomain (CD), cryoID, and ModelAngelo (MA). The first two cases are CATH domains and the other two are SCOPe domains. The resolutions of four cases are 4.02, 5, 6.8, and 8.2 Å. Ø means null.

Model	CATH n=148				SCOPe n=129			
	K=1 (%)	K=5 (%)	K=10 (%)	mAP	K=1 (%)	K=5 (%)	K=10 (%)	mAP
w/o Residual U-Net branch	2.03 (3✓)	31.76 (47✓)	43.92 (65✓)	0.08	32.56 (42✓)	55.81 (72✓)	63.57 (82✓)	0.34
w/o Swin-Conv U-Net branch	6.08 (9✓)	37.16 (55✓)	44.59 (66✓)	0.11	39.53 (51✓)	55.81 (72✓)	68.22 (88✓)	0.42
w/o self-supervised training	2.03 (3✓)	7.43 (11✓)	12.16 (18✓)	0.02	0.00 (0✓)	0.00 (0✓)	6.20 (8✓)	0.01
w/o alignment training	0.00 (0✓)	3.38 (5✓)	6.76 (10✓)	0.02	8.53 (11✓)	8.53 (11✓)	8.53 (11✓)	0.01
CryoDomain	**6.76 (10✓)**	**41.22 (61✓)**	**60.14 (89✓)**	**0.15**	**56.59 (73✓)**	**73.64 (95✓)**	**88.37 (114✓)**	**0.47**

Table 2: Top-K Accuracy (↑) and mAP (↑) on the CATH/SCOPe test sets.

from its use of sequence search, a precise matching retrieval method. When amino acid fragments are well identified, the correct sequences can be searched and found at ease. While a high error rate in fragment recognition leads to no match and null results, producing a low false positive rate. This approach relies on the accuracy of amino acid recognition, which poses significant challenges at low resolution. In contrast, CryoDomain finds all similar embeddings within a given cutoff in a well constructed embedding space. When the distance between embeddings of different domain types is not sufficiently large, CryoDomain may retrieve incorrect domain types, leading to a higher false positive rate than ModelAngelo. Although CryoDomain may have a higher false positive rate, its usability and accuracy on low-resolution maps suggest that it has greater potential for broader application.

Case study. We present the domain identification results of CryoDomain, cryoID, and ModelAngelo on four density maps with various resolutions (Figure 10). cryoID gave wrong protein domains for all the four density maps. ModelAngelo gave the correct domain in the Top-1 for the 4.02 Å density map, the wrong domain for the 5 Å density map, and null results for the others. CryoDomain successfully identified the protein domains in the Top-1 for all four density maps. Furthermore, we employed the *Phenix.dock_in_map* to dock the identified domain atomic structures into the input density maps. We then computed the correlation coefficient (CC_{box} (Afonine et al. 2018)) between the rank 1 atomic structures and the input density maps which reflects the agreement between them. We found that CryoDomain gave protein domain atomic structures with higher CC_{box} than ModelAngelo even when ModelAngelo gave atomic structures with the correct protein domain type, shown in the first case (0.70 v.s. 0.69). And the much lower CC_{box} is consistent with the wrong domain type, as shown in the second case.

Ablation study. To fully assess CryoDomain's components and training, we conducted an ablation study (Table 2). The removal of either the Residual or Swin-Conv U-Net branch from the DensityTower resulted in a decrease in domain identification accuracy. This suggests that both branches are crucial, likely due to their complementary roles in capturing different level features within the density maps. Further, we assessed the impact of training strategies. Omitting self-supervised training led to poorer feature extraction capabilities, as this phase is pivotal for learning robust representations from unlabelled data. Similarly, skipping cross-modal alignment training hindered the model's ability to map density map embeddings to corresponding structural embeddings effectively, as this process benefited from two modalities' information. These results prove that the design of DensityTower and training strategies are effective.

Conclusion

In this paper, we introduce CryoDomain to identify protein domains from low-resolution density maps. Unlike previous works, CryoDomain expands recognition units from amino acids to protein domains, overcoming the difficulty caused by high noise level in low-resolution maps. In addition, CryoDomain innovatively uses a dual-tower network to align density maps and atomic structures in a shared low-dimension space. On two benchmarks constructed from CATH/SCOPe, CryoDomain outperforms previous methods on domain identification from low-resolution maps. This facilitates unbiased discovery of new proteins and cellular landscape by cryo-EM.

Acknowledgments

We thank Associate Professor Ruixue Wan of Westlake University for her expert discussions and valuable guidance. This work is supported by the National Natural Science Foundation of China (Grants No. 32230018 and 32125007 to Q.C.Z., 32400548 to K.X.); the State Key Research Development Program of China (Grant No. 2022YFF1203100 to Q.C.Z.). K.X. was supported by the Beijing Frontier Research Center for Biological Structure (FRCBS) Advanced Innovation Fellowship.

References

Afonine, P. V.; Klaholz, B. P.; Moriarty, N. W.; Poon, B. K.; Sobolev, O. V.; Terwilliger, T. C.; Adams, P. D.; and Urzhumtsev, A. 2018. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D: Structural Biology*, 74(9): 814–840.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; and Bourne, P. E. 2000. The protein data bank. *Nucleic acids research*, 28(1): 235–242.

Buljan, M.; and Bateman, A. 2009. The evolution of protein domain families. *Biochemical Society Transactions*, 37(4): 751–755.

Chang, A.; Xiang, X.; Wang, J.; Lee, C.; Arakhamia, T.; Simjanoska, M.; Wang, C.; Carlomagno, Y.; Zhang, G.; Dhingra, S.; et al. 2022a. Homotypic fibrillization of TMEM106B across diverse neurodegenerative diseases. *Cell*, 185(8): 1346–1355.

Chang, L.; Wang, F.; Connolly, K.; Meng, H.; Su, Z.; Cvirkaite-Krupovic, V.; Krupovic, M.; Egelman, E. H.; and Si, D. 2022b. DeepTracer-ID: De novo protein identification from cryo-EM maps. *Biophysical Journal*, 121(15): 2840–2848.

Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvasy, G.; Mazaré, P.-E.; Lomeli, M.; Hosseini, L.; and Jégou, H. 2024. The Faiss library.

Fox, N. K.; Brenner, S. E.; and Chandonia, J.-M. 2014. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic acids research*, 42(D1): D304–D309.

Fraser, J. S.; and Murcko, M. A. 2024. Structure is beauty, but not always truth. *Cell*, 187(3): 517–520.

Gao, J.; Tong, M.; Lee, C.; Gaertig, J.; Legal, T.; and Bui, K. H. 2023. DomainFit: Identification of Protein Domains in cryo-EM maps at Intermediate Resolution using AlphaFold2-predicted Models. *Biorxiv*.

Greene, L. H.; Lewis, T. E.; Addou, S.; Cuff, A.; Dallman, T.; Dibley, M.; Redfern, O.; Pearl, F.; Nambudiry, R.; Reid, A.; et al. 2007. The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic acids research*, 35(suppl_1): D291–D297.

Gui, M.; Wang, X.; Dutcher, S. K.; Brown, A.; and Zhang, R. 2022. Ciliary central apparatus structure reveals mechanisms of microtubule patterning. *Nature structural & molecular biology*, 29(5): 483–492.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Ho, C.-M.; Li, X.; Lai, M.; Terwilliger, T. C.; Beck, J. R.; Wohlschlegel, J.; Goldberg, D. E.; Fitzpatrick, A. W.; and Zhou, Z. H. 2020. Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nature methods*, 17(1): 79–85.

Jamali, K.; Käll, L.; Zhang, R.; Brown, A.; Kimanius, D.; and Scheres, S. H. 2024. Automated model building and protein identification in cryo-EM maps. *Nature*, 1–2.

Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873): 583–589.

Kühlbrandt, W. 2014. The resolution revolution. *Science*, 343(6178): 1443–1444.

Lawson, C. L.; Patwardhan, A.; Baker, M. L.; Hryc, C.; Garcia, E. S.; Hudson, B. P.; Lagerstedt, I.; Ludtke, S. J.; Pintilie, G.; Sala, R.; et al. 2016. EMDataBank unified data resource for 3DEM. *Nucleic acids research*, 44(D1): D396–D403.

Liao, M.; Cao, E.; Julius, D.; and Cheng, Y. 2013. Structure of the TRPV1 ion channel determined by electron cryomicroscopy. *Nature*, 504(7478): 107–112.

Nogales, E.; and Mahamid, J. 2024. Bridging structural and cell biology with cryo-electron microscopy. *Nature*, 628(8006): 47–56.

Pintilie, G. D.; Zhang, J.; Goddard, T. D.; Chiu, W.; and Gossard, D. C. 2010. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *Journal of structural biology*, 170(3): 427–438.

Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; and Brubaker, M. A. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods*, 14(3): 290–296.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Taheri-Ledari, M.; Zandieh, A.; Shariatpanahi, S. P.; and Eslahchi, C. 2022. Assignment of structural domains in proteins using diffusion kernels on graphs. *BMC bioinformatics*, 23(1): 369.

Taylor, D. W.; Zhu, Y.; Staals, R. H.; Kornfeld, J. E.; Shinkai, A.; van der Oost, J.; Nogales, E.; and Doudna, J. A. 2015. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, 348(6234): 581–585.

Turk, M.; and Baumeister, W. 2020. The promise and the challenges of cryo-electron tomography. *FEBS letters*, 594(20): 3243–3261.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Zhang, K.; Li, Y.; Liang, J.; Cao, J.; Zhang, Y.; Tang, H.; Fan, D.-P.; Timofte, R.; and Gool, L. V. 2023. Practical blind image denoising via Swin-Conv-UNet and data synthesis. *Machine Intelligence Research*, 20(6): 822–836.