

CryoNet.Discovery identifies protein structures from cryo-EM density maps across resolutions through cross-modal recognition

Muzhi Dai^{1,5}, Kui Xu^{2,5}, Yuhao Fei^{3,5}, Yunjian Zhang^{4,5}, Ruiyun Yang¹, Jianlin Lei¹, and Qiangfeng Cliff Zhang^{1,*}

1 State Key Laboratory of Membrane Biology, Beijing Frontier Research Center for Biological Structures, Tsinghua-Peking Joint Center for Life Sciences, Ministry of Education Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Center for Life Sciences and Artificial intelligence, School of Life Sciences, Tsinghua University; Beijing, 100084, China.

2 State Key Laboratory of Membrane Biology, Beijing Tsinghua Institute for Frontier Interdisciplinary Innovation, Beijing, 102202, China.

3 School of Pharmacy, China Pharmaceutical University, Nanjing, 211198, China

4 School of Information Science and Engineering, Lanzhou University, Lanzhou, 730000, China.

5 These authors contributed equally to this work.

*Correspondence: qc Zhang@tsinghua.edu.cn (Q.C.Z.)

Abstract

Cryo-electron microscopy (cryo-EM) is increasingly used not only for structure determination but as a discovery tool for investigating heterogeneous and previously uncharacterized molecular assemblies. Realizing this potential hinges on identifying structures within experimental density maps—a task that grows harder as resolution declines. We developed CryoNet.Discovery, a deep-learning framework that aligns density maps and atomic structures in latent space, recasting structure identification

through cross-modal retrieval. On CATH and SCOPe domain benchmarks, it achieved Top-1 accuracies of 85.7% and 84.4%, respectively, at low resolutions (6–10 Å), outperforming ModelAngelo, cryoID, and DomainFit across a wide resolution range, with the largest gains at low resolutions. Extending from domain- to single-chain identification, it scaled to 12,114 protein chains from 1,488 EMDB maps, recovering the correct fold for 89.3% of queries and revealing structurally conserved yet sequence-divergent relationships missed by sequence-based search. CryoNet.Discovery thus provides a scalable foundation for interpreting unresolved cryo-EM maps and advancing cryo-EM as an engine for structural discovery.

Introduction

Cryo-electron microscopy (cryo-EM) has become an instrumental technology for determining macromolecular structures¹⁻⁵. Much of its development has focused on pushing resolution toward the atomic level, enabling *de novo* model building and detailed mechanistic interpretation^{3,6-9}. Yet the field is rapidly expanding toward larger, heterogeneous, and endogenous assemblies, where experiments frequently capture compositionally complex specimens whose constituents are unknown *a priori*. This makes a new task rate-limiting: identifying which molecular component occupies a given density region in the first place¹⁰⁻¹³. In this setting, cryo-EM is no longer merely a tool for determining the structure of a predefined target, but a discovery instrument capable of revealing unexpected molecular components within complex biological scenes^{14,15}.

This identification task has relied on two broad classes of methods, both of which lose discriminative power as resolution declines. Sequence-assignment methods, including ModelAngelo⁷ and cryoID¹⁶, trace backbone geometry and infer sequence identity from density features. These methods are most effective at near-atomic resolution, where backbone and side-chain features are discernible; their

discriminative power declines as these features blur at intermediate to low resolution. Fitting-based approaches such as DomainFit¹⁷ instead fit candidate structures into the density. Their accuracy depends on the availability of accurate candidate models and their agreement with the target density, making identification sensitive to structural prediction errors and conformational variation. Furthermore, because each candidate must be explicitly fitted and evaluated, exhaustive search is computationally expensive and scales poorly as the number of candidate structures grows. Together, these limitations expose a critical gap: no current method can accurately and efficiently identify structures at intermediate to low resolution, where global shape and topology may be visible but sequence-level features are absent.

We reasoned that this structural identification problem is conceptually analogous to modern face recognition, one of the most successful applications of artificial intelligence. These systems combine exceptional accuracy with highly rapid search across very large identity databases by mapping images into an embedding space in which representations of the same person cluster together despite variations in lighting, pose, or quality^{18,19}. Known individuals are encoded in advance to construct a pre-indexed database, and a new image can be identified efficiently by nearest-neighbor matching. This design transforms recognition from repeated case-by-case comparison into scalable retrieval within a learned representation space.

Here we present CryoNet.Discovery, a deep learning framework that extends and advances this retrieval principle to the identification of protein structures from cryo-EM density maps. Unlike face recognition, however, the query and reference belong to fundamentally different modalities: the query is a density map, whereas the references are atomic structures. CryoNet.Discovery therefore uses a dual-tower architecture to encode each modality separately and learn an alignment between their representation spaces, *i.e.*, a cross-modal correspondence between density and structure. Reference structures are encoded and indexed in advance, allowing a query density to retrieve matching structures by nearest-neighbor search.

CryoNet.Discovery thereby advances from single-modal image retrieval in modern face recognition to cross-modal structural identification for cryo-EM analysis.

We first evaluated CryoNet.Discovery at the domain level, using CATH^{20,21} and SCOPe²² classifications. On benchmarks derived from these resources, CryoNet.Discovery reached Top-1 accuracies of 85.7% (CATH) and 84.4% (SCOPe) at 6–10 Å, surpassing cryoID, ModelAngelo, and DomainFit, with its advantage increasing as resolution declined. Also importantly, by replacing exhaustive candidate fitting with nearest-neighbor search, CryoNet.Discovery accelerates identification by up to two orders of magnitude. We further extended the framework from domain-level to whole-chain identification and screened 12,114 chains from 1,488 maps in the Electron Microscopy Data Bank (EMDB)²³. CryoNet.Discovery recovered the correct fold for 89.3% of queries and revealed structurally conserved yet sequence-divergent relationships missed by sequence-based search. Together, these results establish CryoNet.Discovery as an accurate, efficient, and scalable computational framework for structure identification, supporting cryo-EM's expanding role in biological discovery.

Results

CryoNet.Discovery aligns density maps and atomic structures for structure identification

CryoNet.Discovery formulates structure identification through cross-modal retrieval between cryo-EM density maps and atomic structures. Given a density region, it retrieves the best-matching structure entries from a pre-indexed reference gallery by comparing learned density map and atomic structure embeddings (**Fig. 1A**). CryoNet.Discovery uses a density tower to encode density maps, and an atom tower to encode atomic structures (**Fig. S1A–B**), with the two embedding spaces aligned through training on known density–structure pairs to establish a correspondence

between the modalities. At inference, it embeds a query density map with the same density tower, and matches it against a precomputed gallery of structure embeddings, retrieving the most likely reference identities by nearest-neighbor search.

The two encoders have distinct architectures tailored to their respective inputs. The density tower passes a three-dimensional (3D) density volume through two parallel convolutional projection layers, whose outputs are passed to a dual-branch 3D U-Net²⁴ comprising a Residual²⁵ U-Net branch and a Swin-Conv U-Net²⁶ branch. The resulting feature map drives two heads: a reconstruction head that predicts the input density volume, and a compression head that applies 3D convolutions²⁷, global pooling²⁸, and fully connected layers to generate a 128-dimensional density map embedding (Z_d). The atom tower accepts domain-level atomic structures and builds residue-level single and pair representations through an atom encoder, a geometry-refinement structure module, and a fusion module, which combines these representations via normalization²⁹, attention, and feed-forward transformations³⁰ to produce a 128-dimensional structure embedding (Z_a).

Once both towers are trained (see below), CryoNet.Discovery operates in two steps that mirror standard retrieval systems: offline indexing and online querying. In the offline indexing step, the atom tower encodes a gallery of reference atomic structures into a queryable, density-aware structure embedding database (DateDB) (**Fig. S1C**). Because only structures present in this gallery can ever be retrieved, its composition defines the space of identifiable answers for a given application (See "Anchor-and-align training and cross-modal retrieval" below). In the online querying step, a single density tower forward pass converts an input density region into a query embedding Z_d , which is matched against DateDB by FAISS inner-product nearest-neighbor search³¹ to return a list of candidate structures ranked by similarity (**Fig. S1D**). This design decouples the computational cost of encoding the reference gallery from query-time identification, enabling rapid search across large structure collections.

Construction of training and benchmark datasets

We first implemented and evaluated CryoNet.Discovery at the domain level, where curated structural classifications provide well-defined identities for structure identification. To train and evaluate CryoNet.Discovery, we assembled three datasets from public repositories: a density map dataset (DMD) for training the density tower, an atomic-structure dataset (ASD) for training the atom tower, and a paired structure–map dataset (PSMD) for aligning the embeddings produced by the two towers. Dataset construction proceeds in three steps: (i) raw data collection and quality control, (ii) domain segmentation and annotation, and (iii) splitting the paired data into a benchmark and augmenting the remaining data for training.

Step 1 — Raw data collection and quality control. We queried the EMDB for all maps deposited on or before September 1, 2025 and retained those associated with a corresponding protein structure in the Protein Data Bank (PDB)³², yielding 21,228 maps (1–20 Å). To ensure consistency between deposited maps and their models, mismatched pairs were removed by an automated correlation filter ($CC_{\text{box}} < 0.2$)³³ followed by manual inspection in ChimeraX³⁴, leaving 17,150 quality-controlled paired maps (corresponding to 17,012 structural models; the number of maps exceeds the number of structures because one structure can be associated with multiple maps).

Step 2 — Domain segmentation and annotation. Each quality-controlled structure was decomposed into its constituent domains, and every structural domain was paired with the corresponding sub-region of its deposited map. Among the 17,012 structural models, 1,098 PDB structures carrying CATH or SCOPe annotations were segmented by these curated domain-type labels, yielding 8,822 CATH-labeled structural domains paired to 8,943 map domains and 1,961 SCOPe-labeled structural domains paired to 2,012 map domains (**Fig. S2A**). The remaining 15,914 unannotated structures were segmented by KluDo³⁵, a structure-based domain segmentation algorithm, yielding 239,375 unannotated structural domains paired to 241,166 map

domains—unlabeled but usable during training (**Fig. S2A**). Together, these annotated and KluDo-segmented pairs constituted the full PSMD: 252,121 map domains matched to 250,158 distinct structural domains.

Step 3 — Training/benchmark split and training data augmentation. The benchmark was drawn from the pairs with CATH/SCOPe annotations by stratified sampling, selecting one representative sample for each domain type within each 1 Å resolution bin to mitigate domain-type bias. This yielded intermediate- to high-resolution (1–6 Å) and low-resolution (6–10 Å) benchmark subsets, comprising 318 and 21 pairs for CATH, and 85 and 77 pairs for SCOPe (no pairs survived filtering in the 10–20 Å range; **Fig. 2A**). To prevent data leakage, any training domain sharing >90% sequence identity with a benchmark domain was removed using MMseqs2^{36,37}, leaving 183,925 paired domains (1–20 Å) for alignment training (**Fig. 2A**).

The two towers are pretrained separately before alignment. The density tower is pretrained on the DMD, which comprises all 183,925 quality-controlled map domains in the PSMD. The atom tower is pretrained on the ASD, which, to broaden domain-type coverage, combines the PSMD's 182,576 structural domains with ~400,000 CATH- and ~210,000 SCOPe-annotated domains derived from a broader collection of PDB structures determined by X-ray³⁸ and other experimental methods. In total, the ASD contains ~790,000 structural domains.

Notably, to promote resolution-invariant embeddings, we apply additional data augmentation at each training epoch: for the PSMD, 10% (18,392) of experimental maps are randomly selected and low-pass filtered to every integer resolution coarser than the deposited value up to 10 Å, and 18,392 randomly selected annotated but unpaired ASD structures are converted into synthetic density maps simulated with *molmap*³⁴ at 1–10 Å.

Anchor-and-align training and cross-modal retrieval

CryoNet.Discovery was trained with a progressive three-stage strategy that separates modality-specific representation learning from cross-modal alignment (**Fig. S1E**). In Stage 1, the density tower and the atom tower were pretrained independently on the DMD and ASD (self-supervised pretraining), respectively, using modality-specific reconstruction objectives³⁹. The density tower was optimized by reconstructing its input density with an MSE loss (L_{density}), encouraging the density map embedding to preserve reproducible shape and secondary-structure features. The atom tower was optimized with a composite structure-supervision loss (L_{atom}) that integrates frame aligned point error (FAPE)⁴⁰, backbone torsion angle, distogram, and predicted local distance difference test (pLDDT) terms, so that its representation retained residue-level geometry and inter-residue relationships.

In Stage 2, contrastive learning⁴¹ was applied to the atom tower using domain-type annotations via a contrastive loss (L_{contrast}) based on cosine similarity, pulling embeddings of the same domain type together and pushing those of different types apart to organize the atomic structure embedding space by domain topology. In Stage 3, the atom tower was frozen, and the density tower was fine-tuned on paired density-structure samples with a cosine-distance alignment loss ($L_{\text{align}} = 1 - \cos(Z_d, Z_a)$). This procedure anchored density map embeddings to the pretrained atomic structure embedding space. This anchor-and-align design minimizes reliance on scarce annotated density-structure pairs by deriving most of the embedding-space structure from the abundant atomic-structure-only set, and yields a final retrieval space that is aligned across modalities. The resulting framework differs from conventional single-modal retrieval: rather than comparing two observations encoded by the same network, it learns a correspondence between two physically distinct modalities and enables one to retrieve the other.

Following the completion of cross-modal training of the two towers, CryoNet.Discovery operates as a retrieval system in two steps—offline database

construction and online query—as introduced above. In the offline step, a gallery of atomic structures is pre-assembled and each member is passed through the frozen atom tower to construct the DateDB. In the online query step, each density map is encoded by the trained density tower and searched against DateDB by FAISS inner-product nearest-neighbor retrieval. Hit structures are then grouped by domain type, retaining the highest-scoring representative of each type to produce a nonredundant Top- k ranking.

CryoNet.Discovery identifies domain structures accurately and efficiently across resolutions

We benchmarked CryoNet.Discovery using the annotation-specific reference galleries: DateDB-CATH (8,822 domains spanning 164 CATH domain types) and DateDB-SCOPe (1,961 domains spanning 116 SCOPe domain types) (**Fig. 2A**). To better approximate real-world identification of previously unseen structures, any DateDB entry sharing more than 90% sequence identity with the training set was excluded from the reference gallery (using MMseqs2). Each query map was then searched against its corresponding DateDB: CATH queries against DateDB-CATH and SCOPe queries against DateDB-SCOPe. A retrieval was considered as correct when the ground-truth CATH or SCOPe domain-type annotation appeared among the Top- k ranked domain types. Performance was reported using Top-1/Top-5 accuracy^{42,43} and the Rank Index, which records the rank position of the correct domain type in the retrieved list.

CryoNet.Discovery achieved the highest Top-1 accuracy among the evaluated methods in each benchmark setting (**Fig. 2B**). At 1–6 Å, it achieved Top-1/Top-5 accuracies of 87.1%/96.5% on CATH and 83.5%/97.6% on SCOPe. At 6–10 Å, performance remained high, reaching 85.7%/100.0% on CATH and 84.4%/94.8% on SCOPe. Notably, Top-1 accuracy was largely preserved from the 1–6 Å to the 6–10 Å resolution bins, with only a slight decrease on CATH, indicating

the robustness of CryoNet.Discovery to the loss of high-resolution density features. Across the four benchmark settings, the strongest baseline lagged CryoNet.Discovery by 5.9 to 42.8 percentage points in Top-1 accuracy, with the performance gap increasing at low resolution. DomainFit and cryoID consistently showed lower Top-1 accuracies across both resolution ranges, whereas ModelAngelo did not return a valid or correct domain-type assignment at 6–10 Å. CryoNet.Discovery also ranked the correct identity near the top of the returned candidates, with mean Rank Indices of 2.4 on both CATH and SCOPe at 1–6 Å, and 1.4 on CATH and 4.7 on SCOPe at 6–10 Å (**Fig. 2C**). By comparison, DomainFit and cryoID showed higher Rank Indices, and ModelAngelo degraded sharply at low resolution (**Fig. S2B**).

Performance varied with domain morphology (**Fig. 2D**). CryoNet.Discovery was most reliable for compact or topologically distinctive folds, including WD40-repeat and ribosomal protein domains, but was less reliable for elongated, repetitive, or coiled-coil-like architectures. DomainFit showed a partially complementary preference, performing better on some elongated folds. These results indicate that the two approaches exploit different structural signals and that low-resolution density envelopes may provide intrinsically insufficient discriminative information for some repetitive architectures.

Beyond high accuracy, CryoNet.Discovery was also the fastest among the evaluated methods across density maps of varying volume, running 2- to 100-fold faster than ModelAngelo, cryoID, and DomainFit (**Fig. 2E**). This acceleration gain follows directly from the retrieval architecture: reference structures are encoded only once during offline database construction, whereas each new query requires only a density tower forward pass and nearest-neighbor search rather than iterative structure fitting or *de novo* model construction.

Representative examples illustrated the performance of CryoNet.Discovery across structurally diverse low-resolution density regions (**Fig. 2F–G and Fig. S2C–D**). Across four density regions spanning compact YVTN

repeat-like/quinoprotein amine dehydrogenase domain, G protein, zinc/RING-finger domain, and F1F0 ATP synthase subunit C, at resolutions ranging from 6.9 to 10.0 Å, CryoNet.Discovery identified the correct domain type at Top-1 rank. Subsequent rigid-body fitting of the identified structures into the corresponding density regions showed close map–model agreement³³ (e.g. CC_{box} values ranging from 0.62–0.81). By contrast, competing methods either identified a different fold or returned no valid candidate in several cases. These examples support the quantitative benchmark by showing that the identified domain structures are geometrically compatible with structurally diverse low-resolution densities. Performance on 2–5 Å maps remained competitive with ModelAngelo and DomainFit (**Fig. S3**), indicating that the low-resolution advantage was not accompanied by a substantial loss at higher resolution.

Together, these results show that CryoNet.Discovery combines accurate and efficient structure identification, with its largest advantage emerging in the low-resolution regime where residue-level structural information is vague.

CryoNet.Discovery maintains robust structure identification across structural classification levels and cryo-EM resolutions

Protein-domain classifications are inherently hierarchical. CATH categorizes domains into Superfamily, Topology, Architecture, and Class, whereas SCOPe classifies domains into Family, Superfamily, Fold, and Class (ordered here from finer to coarser). This hierarchy provides a useful framework for evaluating structure identification across progressively broader structural features: although family-level discrimination represents the most demanding task, a robust method should always recover the correct higher-level structural classification when fine structural features are degraded. We therefore evaluated Top-1 accuracy across all hierarchical levels of the CATH and SCOPe classification systems to determine whether identification errors at the finest level remained topologically meaningful rather than arbitrary.

For intermediate- to high-resolution benchmarks, CryoNet.Discovery, DomainFit, and ModelAngelo all achieved strong performance across classification levels (**Fig. 3A**). CryoNet.Discovery achieved Top-1 accuracies of 87.1%/83.5% at the finest CATH Superfamily and SCOPe Family levels, increasing to 97.8%/94.1% at the coarsest Class level on the two benchmarks. ModelAngelo remained competitive (79.9%/77.6% across all levels), whereas DomainFit was weaker at finer levels (66.7%/52.9% at Superfamily/Family; 82.7%/65.9% at Class), and cryoID performed substantially worse (8.2%/34.1% at Superfamily/Family; 38.7%/50.6% at Class). This superiority extended to the low-resolution CATH benchmark, where CryoNet.Discovery remained uniformly accurate, achieving Top-1 accuracies of 85.7%, 85.7%, 95.2%, and 95.2% at the Superfamily, Topology, Architecture, and Class levels, respectively. By contrast, DomainFit stayed modest across the hierarchy (66.7% at Superfamily level) and improved only slightly toward the lower levels (71.4% at the Class level), while cryoID essentially failed (4.8% across the three finer levels), and ModelAngelo yielded no valid results at any level in this resolution range. The SCOPe benchmark showed the same pattern, with CryoNet.Discovery maintaining high accuracy from the Family (84.4%) to the Class (90.9%) level. Across both benchmarks, all methods showed modest improvements toward coarser classification levels, reflecting the reduced difficulty of broader structural assignments. Nevertheless, CryoNet.Discovery consistently maintained the highest accuracy across every level of the CATH and SCOPe hierarchies.

To isolate this resolution dependence directly, we curated three EMPIAR datasets^{44,45}, each containing raw two-dimensional (2D) particle images paired with a corresponding PDB atomic model. From each complete particle set, we randomly subsampled particles and *de novo* reconstructed 3D maps using cryoSPARC⁴⁶, yielding four maps of the same molecular complex spanning global resolutions from 6 to 10 Å (**Table S1**). Eight CATH-annotated domains were extracted across these reconstructed maps, establishing a benchmark in which performance differences could be attributed primarily to resolution.

Across the four resolution intervals, CryoNet.Discovery consistently maintained low Rank Indices (**Fig. 3B**), indicating that the correct domain type remained highly ranked among the candidates even as resolution decreased. In contrast, DomainFit deteriorated progressively as resolution decreased, whereas cryoID and ModelAngelo did not return valid assignments at 6–7 Å or lower resolutions, yielding the maximum Rank Index of 164. Even at 9–10 Å, CryoNet.Discovery correctly identified the Top-1 domain for most query regions, and subsequent rigid-body fitting of these atomic models into the experimental density confirmed close spatial agreement (**Fig. 3C and Fig. S4**), demonstrating that the identified structures were not merely correct in domain type but physically compatible with the density. Together, these results show that CryoNet.Discovery preserves accurate domain-level structure identification as resolution decreases, whereas both sequence-alignment and density-fitting methods progressively lose discriminative power or fail entirely under the same conditions.

Dual-branch encoding and cross-modal alignment are both essential to CryoNet.Discovery's performance

To dissect the contributions of each component, we performed leave-one-out ablation studies starting from the full CryoNet.Discovery model, removing a single architectural module, training stage, or data augmentation strategy while keeping all other settings fixed and retraining the model to convergence (**Fig. 3D**). The full CryoNet.Discovery model provided the reference baseline, achieving Top-1 accuracies of 87.0% on CATH and 84.0% on SCOPe, with corresponding mean average precision (mAP)⁴⁷ values of 0.75 and 0.84, respectively.

We first tested the two branches of the density tower architecture. Removing the Residual U-Net branch reduced Top-1 accuracy to 70.2% on CATH and 56.8% on SCOPe, with mAP values dropping to 0.65 and 0.61. Removing the Swin-Conv U-Net branch caused a substantially larger drop, reducing Top-1 accuracy to 10.0% on CATH and 9.3% on SCOPe, with mAP values of only 0.16 and 0.18. These results

show that both branches contribute to identification performance: the Residual U-Net branch provides complementary density features, whereas the Swin-Conv U-Net branch contributes the dominant discriminative signal, and its removal accordingly produces the sharper performance collapse.

Beyond architectural design, we next examined the contribution of the three-stage training strategy. Disabling modality-specific self-supervised pretraining led to a consistent but moderate reduction in performance, decreasing Top-1 accuracy to 79.9% on CATH and 74.1% on SCOPe, with mAP values of 0.74 and 0.75, indicating that pretraining refines rather than enables identification. In contrast, removing the cross-modal embedding framework abolished identification performance almost entirely, reducing Top-1 accuracy to 0.3% on CATH and 1.2% on SCOPe, with mAP values of 0.02 and 0.03, respectively. Thus, the defining capability of CryoNet.Discovery does not arise from either encoder alone, but from the learned correspondence between density maps and atomic structure that enables cross-modal comparison.

Finally, we evaluated the two data augmentation strategies used during training. Removing *molmap* data augmentation reduced Top-1 accuracy to 80.5% on CATH and 71.6% on SCOPe, with mAP values of 0.72 and 0.73. Removing low-pass filtering data augmentation led to a larger reduction, especially on SCOPe, lowering Top-1 accuracy to 79.4% on CATH and 67.9% on SCOPe, with mAP values of 0.70 and 0.72. These results indicate that both augmentation strategies improve the robustness of density representation learning, with *molmap* data augmentation expanding the diversity of simulated density inputs and low-pass filtering data augmentation improving tolerance to resolution degradation. Overall, the ablation studies show that CryoNet.Discovery derives its performance from the combination of dual-branch density encoding, cross-modal alignment, and resolution-aware density augmentation. Among these components, cross-modal alignment is the defining requirement for density-to-structure identification, whereas

the encoder architecture and data augmentation largely determine its accuracy and robustness to resolution loss.

CryoNet.Discovery enables *de novo* structural annotation of experimental maps across both single-particle cryo-EM and cryo-ET

We next evaluated CryoNet.Discovery on experimental density maps lacking domain annotation: one single-particle analysis (SPA) cryo-EM map⁴⁸ and one subtomogram-averaging cryo-electron tomography (cryo-ET) map^{49,50}. In both cases, we manually segmented the density into visually discrete, self-contained regions (representing putative independently folded structural units) and queried them against DateDB-CATH-full (5,689 CATH domain types).

For the SPA regime, we examined EMD-8317, a 7.9 Å map of wild-type *E. coli* GroEL⁵¹ derived from the apo reference state of the PepQ-bound study and without an assigned atomic structure model (**Fig. 4A**). We manually segmented a representative subunit of the D7-symmetric assembly into three density regions, and independently queried each of them against DateDB-CATH-full. CryoNet.Discovery returned a ranked list of CATH superfamilies for each region, with the top five candidates shown in **Fig. 4B**. The Top-1 structure showed consistent agreement with their corresponding densities after rigid-body fitting ($CC_{\text{box}} = 0.57\text{--}0.72$; **Fig. 4C**). Extending the analysis to all subunits revealed reproducible domain assignments across symmetry-related positions (**Fig. 4D**). The resulting model recovered the established three-domain architecture of GroEL determined by crystallography⁵² and agreed with the high-resolution crystal model used to initialize the reconstruction, which was derived from a GroEL-DprE1 crystallization study⁵³ (**Fig. S6A**), showing no systematic mismatch with the corresponding density regions (**Fig. S6B**).

Cryo-ET poses a more stringent test because of its lower signal-to-noise (SNR) ratio and missing-wedge artefacts. We examined EMD-12454, a 6.2 Å subtomogram-averaging reconstruction lacking a structure model⁵⁴ (**Fig. 4E**). We first

manually segmented the C6-symmetrized map into twelve discrete density regions, comprising two putative independently folded units in each of six repeated protomers, and selected one representative of each unit to query against DateDB-CATH-full. The Top-1 assignments showed strong agreement with their corresponding densities after rigid-body fitting (**Fig. 4F–H**), where the HIV-1 capsid N-terminal domain (CATH ID: 1.10.375.10) was assigned to the inner-ring region and the retroviral capsid C-terminal domain (CATH ID: 1.10.1200.30) was assigned to the outer-ring region. The resulting model again agreed closely with the recently established architecture of the hexameric HIV-1 capsid lattice⁵⁵ (**Fig. S6C**), with no systematic mismatch with the corresponding density regions (**Fig. S6D**). Together, these examples show that CryoNet.Discovery can recover biologically coherent domain-level organization from both SPA cryo-EM and cryo-ET maps without sequence input, supporting reference-based annotation of density maps that currently lack atomic models.

CryoNet.Discovery generalizes to whole-chain structure identification and recovers structurally similar, sequence-divergent proteins across the EMDB

We next evaluated whether CryoNet.Discovery framework could support whole-chain structure identification (**Fig. 5A**). This objective differs from the domain-level benchmark because query regions may contain multiple domains, flexible linkers, and substantially greater variation in molecular size, reflecting how the method is used in practice, where a query arrives as a complete protein chain rather than as a pre-segmented set of domains.

We collected all EMDB entries deposited on or after January 1, 2026 as the query dataset. As EMDB entries are typically multi-chain assemblies, we split each map into per-chain density regions, yielding 12,114 single-chain queries from 1,488 EMDB entries. The reference database, DateDB-PDB50-clustered, was built by clustering all PDB chains at 50% sequence identity and keeping one representative per cluster, each encoded by the atom tower and indexed for similarity search. For each query, we embedded the density with the density tower and ranked all candidates

in DateDB-PDB50-clustered based on embedding similarity to the query embedding. The highest-ranking candidates were subsequently fitted into the density and scored using map–model correlation (CC_{box}), after which the five best-correlating clusters were retained. We then expanded each retained cluster to all member chains, aligned every member to the deposited ("ground-truth") sequence with EMBOSS Needle⁵⁶ to obtain the cluster's highest sequence identity, and structurally aligned the corresponding highest-similarity member to the deposited query model using USalign to calculate TM-score^{57,58}.

Overall, 89.3% of queries yielded at least one candidate, among the five candidate clusters, with a TM-score of at least 0.5, a commonly used threshold for same-fold structural similarity, and 74.8% reached a TM-score of at least 0.7, indicating a more stringent level of structural agreement⁵⁹ (**Fig. S7A**). Identification performance was resolution-dependent but structurally robust (**Fig. 5B**), mean TM-scores declined from 0.75–0.95 below 5 Å to 0.65–0.74 beyond 5 Å, consistent with progressive loss of local atomic detail. Notably, sequence similarity declined more sharply, with only 52.6% of queries a candidate with sequence identity ≥ 0.3 and 37.6% with sequence identity ≥ 0.9 (**Fig. S7A**). This divergence reflects a basic principle of structural biology: as map resolution decreases, sequence becomes increasingly difficult to recognize, whereas the more conserved structural fold remains accessible.

This resulted in a set of structurally similar yet highly sequence-divergent protein pairs^{60,61}. In the EMD-73999 (3.6 Å; **Fig. 5C**), the deposited Pakpunavirus P7-1 neck protein (*PDB-9ZBS_i*)⁶² matched the bacteriophage Sf14 neck protein (*PDB-9C39_F*)⁶³ at only 5% sequence identity yet with clear structural and density agreement (TM-score 0.78, RMSD 2.88 Å). In the EMD-56117 map⁶⁴ (6.4 Å; **Fig. 5D**), the target of rapamycin complex subunit LST8 protein (*PDB-9TPW_C*) matched a chain of the yeast 48S preinitiation complex⁶⁵ (*PDB-3JAP_q*) also at 5% sequence identity yet with clear structural agreement (TM-score 0.87, RMSD 2.55 Å). Such

pairs are intriguing in their own right and may warrant further investigation into their structural and evolutionary relationships.

Together, these results show that CryoNet.Discovery supports repository-scale whole-chain identification and can enrich for structurally related proteins with limited sequence similarity.

Discussion

CryoNet.Discovery establishes a cross-modal AI framework for identifying protein structures directly from cryo-EM density. Inspired by the retrieval principle underlying modern face recognition, it advances recognition from comparisons within a single data modality to correspondence between two physically distinct representations: density maps and atomic structures. Once the structure gallery is encoded and indexed, each query requires only density map embedding and nearest-neighbor search. This design enabled accurate domain-level identification across 1–10 Å, accelerated identification by up to two orders of magnitude, and supported both experimental-map annotation and repository-scale screening.

CryoNet.Discovery operates at the level of information supported by experimental density. Intermediate- and low-resolution maps may lose residue-level features while retaining secondary-structure organization, domain topology, and overall shape. The model learns to connect these density features with atomic structures in embedding space rather than directly reconstructing atomic detail not resolved by the map. The ablation results support this design principle: cross-modal alignment was essential for identification, the two density-encoding branches contributed complementary structural features, and resolution-aware augmentation improved robustness to resolution degradation.

Several limitations define the practical scope of CryoNet.Discovery. First, the specificity of identification is fundamentally limited by the information content of the density map. When multiple proteins share comparable global folds or domain topologies, the map may support several structurally plausible candidates rather than a unique identity. In such cases, a ranked candidate list accompanied by calibrated confidence estimates can provide testable hypotheses for subsequent biological interpretation and experimental validation. This limitation is reflected in our finding that compact and topologically distinctive domains were identified more reliably than elongated, repetitive, or coiled-coil architectures, as these structures can produce similar low-resolution envelopes. Second, CryoNet.Discovery identification is constrained by the coverage of the reference gallery. Because nearest-neighbor search always returns a candidate, the current framework cannot reliably determine when an appropriate structure is absent from the database; confidence calibration, abstention criteria, and out-of-distribution detection will therefore be important for preventing unsupported assignments. Third, performance depends on accurate segmentation of the query density. Boundary errors, overlapping densities, conformational heterogeneity, flexible regions, and missing-wedge effects in cryo-ET may propagate into identification. Finally, biases in deposited training data, including those related to protein-family representation, molecular size, map processing, and experimental quality, may limit performance and generalization to underrepresented structures and map conditions.

In sum, CryoNet.Discovery establishes a new AI framework for structural recognition by aligning cryo-EM maps and atomic structures for accurate, efficient, and sequence-independent structure identification. Its pre-indexed architecture enables recognition to scale across large collections of experimental maps while remaining effective across resolutions. Beyond the applications demonstrated here, we envision CryoNet.Discovery becoming a broadly applicable component of future cryo-EM and cryo-ET analysis studies spanning increasingly complex native

assemblies, cellular environments, and structural atlases, thereby supporting exploration and hypothesis generation in discovery-oriented structural biology.

Methods

1. Network architecture and cross-modal retrieval

1.1 Architecture overview

CryoNet.Discovery is a dual-tower neural network that embeds protein domain density maps and atomic structures into an aligned embedding space. It consists of a density encoder (Density tower) and an atomic structure encoder (Atom tower), each of which encodes its respective modality input into a 128-dimensional embedding for subsequent cross-modal retrieval.

1.2 Density tower architecture

The density tower takes a three-dimensional cryo-EM density map as input. The input density volume is first transformed by two parallel convolutional projection layers. The resulting features are combined by element-wise addition and then fed into a dual-branch U-Net composed of a Residual²⁵ U-Net²⁴ branch and a Swin-Conv U-Net²⁶ branch (**Fig. S1A**).

The backbone comprises two parallel branches that share the same four-level multiscale encoder–decoder layout. Spatial resolution is reduced by a factor of two at each encoder level and restored symmetrically during decoding, ensuring alignment between branches at all scales.

1.3 Atom tower architecture

The atom tower consists of an atomic encoding module (Atom encoder), a structure module, and a fusion module (**Fig. S1B**). The network takes domain-level atomic structures as input and uses the atom encoder to generate a single representation and a pair representation in parallel. These representations are provided in parallel to the structure module for geometric refinement and to the fusion module for embedding

generation, with the fusion module outputting a 128-dimensional atomic structure embedding (Z_a).

1.4 Construction of the density-aware structure embedding database (DateDB)

After cross-modal alignment, we constructed a density-aware structure embedding database (DateDB) to enable protein domain identification at inference time (**Fig. S1C**). To build DateDB, we used the trained atom tower to encode all collected protein domain atomic structures and stored the resulting atomic structure embeddings (Z_a).

1.5 Retrieval with CryoNet.Discovery

After density and atomic structure representations are aligned, CryoNet.Discovery identifies protein domains by comparing density map and atomic structure embeddings in the aligned embedding space (**Fig. S1D**). An input density region is encoded by the density tower to obtain a density map embedding (Z_d), which is used as a query to search the DateDB containing atomic structure embeddings (Z_a) generated by the atom tower.

2. Dataset preparation and pre-processing

2.1 Dataset preparation

We constructed a dataset spanning a wide resolution range (1–20 Å), with diverse density qualities and structural types. The raw cryo-EM density maps were sourced from the EMDB (deposited on or before September 1, 2025) and paired with atomic structures from the PDB through cross-referencing information.

For domain-level analysis, atomic structures corresponding to cryo-EM density maps were partitioned into domains using annotations from the CATH and SCOPe databases. Density regions corresponding to annotated domains were extracted from cryo-EM density maps according to atomic coordinates. Automatic quality control

and manual inspection were then applied to remove misaligned or anomalous samples, including cases with poor density–model correspondence or ambiguous domain-to-density assignment. Because one deposited atomic model can correspond to multiple EMDB maps, the curated annotated set contained 8,943 CATH and 2,012 SCOPe-labeled density–atomic structure pairs, corresponding to 8,822 and 1,961 domain atomic structures, respectively.

Because many atomic structures corresponding to cryo-EM density maps lacked CATH or SCOPe domain annotations, KluDo, a structure-based domain segmentation tool, was further used to predict domain boundaries for these atomic structures. KluDo was used only to define domain boundaries and did not assign domain-type labels or classifications. Applying the same coordinate-based density extraction and quality-control procedure to KluDo-defined domains yielded 241,166 unlabeled domain density–atomic structure pairs.

Given the limited number of labeled density–atomic structure pairs, the benchmarks were prioritized and constructed exclusively from pairs with reliable CATH or SCOPe annotations, and stratified sampling was applied to mitigate domain-type bias, retaining only one sample per domain type within each 1 Å resolution bin. The resulting density maps were divided into intermediate- to high-resolution (1–6 Å) and low-resolution (6–20 Å) benchmark subsets. After domain annotation and benchmark filtering, no eligible CATH or SCOPe entries above 10 Å remained. Thus, the retained low-resolution benchmarks covered 6–10 Å and comprised 21 CATH maps and 77 SCOPe maps, while the intermediate- to high-resolution benchmarks (1–6 Å) comprised 318 CATH maps and 85 SCOPe maps.

The training set was constructed from the remaining density-atomic structure pairs, after excluding any training domains that shared more than 90% sequence identity with domains in the benchmarks, thereby ensuring dataset independence and

avoiding information leakage. The final training set contained 183,925 density-atomic structure pairs, covering a broad range of structural types and resolutions (1–20 Å).

2.2 Pre-processing

During pre-processing, density maps and atomic structures are standardized to ensure consistency in numerical scale, spatial representation, and geometric description.

2.3 DateDB construction

To construct benchmark DateDBs for the CATH and SCOPe evaluations, we used the annotated atomic structures associated with each benchmark. DateDB-CATH was built from 8,822 annotated atomic structures covering 164 CATH domain types, whereas DateDB-SCOPe was built from 1,961 annotated atomic structures covering 116 SCOPe domain types. Each atomic structure was encoded using CryoNet.Discovery’s atom tower, and the resulting atomic structure embeddings were stored together with their domain annotations and metadata for subsequent density-to-structure retrieval. For each test density map, the corresponding benchmark-specific DateDB was ensured to contain the ground-truth domain type, enabling domain-type-level identification evaluation.

To support broader identification analyses, we further constructed full-scale DateDBs from ~400,000 CATH-annotated and ~210,000 SCOPe-annotated atomic domains. To remove exact sequence redundancy, the collected domains were clustered at 100% sequence identity using the MMseqs2 *easy-linclust* tool, and one representative was retained from each cluster. The resulting non-redundant databases, DateDB-CATH-full and DateDB-SCOPe-full, covered 5,689 CATH domain types and 4,045 SCOPe domain types, respectively, thereby providing broad structural-category coverage for domain-type semantic representation and identification analysis.

3. Training strategy and loss functions

3.1 Training strategy

CryoNet.Discovery is trained using a three-stage strategy that separates modality-specific representation learning from cross-modal alignment. The training procedure consists of (i) modality-internal self-supervised learning, (ii) contrastive learning on the atomic structure modality, and (iii) cross-modal alignment between density and atomic structure embeddings (**Fig. S1E**).

3.2 Loss functions and training settings

CryoNet.Discovery is trained using a multi-stage optimization scheme, with different loss functions applied at different stages to learn modality-specific representations and to align cryo-EM density map embeddings with atomic structure embeddings.

4. Evaluation

4.1 Retrieval evaluation metrics

To evaluate the performance of CryoNet.Discovery in domain-level structure identification, we use two primary metrics, Top-k Accuracy and Rank Index.

4.2 Runtime benchmarking

Runtime was measured on the same server using identical query density maps and candidate databases for each method. For CryoNet.Discovery, the reported runtime included density encoding and nearest-neighbor retrieval from precomputed DateDB embeddings; for comparison methods, the reported runtime included the complete domain-identification workflow under the benchmarking settings used in this study.

4.3 Case-study fitting and map-model correlation assessment

For representative case studies, the Top-1 retrieved atomic structure was rigid-body fitted into the corresponding segmented density region before calculating map-model correlation coefficient metrics, including CC_{box} , CC_{mask} , CC_{volume} , and CC_{peaks} . The

same fitting and scoring procedure was applied to comparison methods whenever valid candidate models were returned.

4.4 Computing environment

All evaluations were performed on a single server equipped with two Intel Xeon Gold 6240 18-core CPUs, an NVIDIA A800-PCIE-80G GPU, and 256 GB of CPU RAM.

5. EMDB-scale structure identification

5.1 EMDB-wide pipeline

We constructed an automated pipeline for large-scale processing of EMDB density maps by extending domain-level structure identification to single-chain level. Structural retrieval was performed against a precomputed embedding database (DateDB-PDB50-clustered).

5.2 Evaluation

Sequence identity's command is:

```
needle -asequence ground_truth_chain.fasta -bsequence candidate_chain.fasta  
-gapopen 10 -gapextend 0.5 -outfile output.needle-auto
```

TM-score's command is:

```
USalign ground_truth_chain.pdb selected_cluster_member.pdb -mm 5
```

Web server

The CryoNet.Discovery web server is freely available for academic use at <https://cryonet.ai/discovery>.

Supplementary Table 1

EMPIAR	Subset	Total particles	Subsampled particles	Symmetry	Resolution (Å)
EMPIAR-10786 ⁶⁶	1	288,659	9,486	C1	6.7
			6,946		7.9
			8,525		8.0
			6,260		9.0
	3	122,220	3,032	C1	6.0
			3,695		7.1
			2,255		8.8
			2,540		9.4
EMPIA R-10005 ⁶⁷	1	35,645	1,542	C4	6.6
			1,369		7.9
			1,110		8.4
			1,251		9.0

Acknowledgments

We thank Ruixue Wan for helpful comments and advice. Q.C.Z. was supported by Science & Technology Fundamental Resources Investigation Program (2025FY100300), National Natural Science Foundation of China (32125007, 32230018), the Xplorer Prize (2024), and Tsinghua University Dushi Program and the Qiantang Professorship. K.X. was supported by the Young Scientists Fund of the National Natural Science Foundation of China (32400548). Y.F. was supported by the Xingyao Leading Scholar Program of China Pharmaceutical University. J.L. was

supported by the National Key Research and Development Program of China (2024YFA1307303).

Author contributions

Q.C.Z. supervised and K.X. conceived the project. K.X. and M.D. designed and implemented the CryoNet.Discovery model. M.D., K.X., Y.F., Y.Z., and J.L. validated the CryoNet.Discovery model and analyzed the results. Q.C.Z., K.X., M.D., Y.F., and R.Y. wrote the manuscript, with input from all the authors.

Declaration of interests

The authors declare no competing interests.

References

- 1 Nogales, E. & Scheres, S. H. Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell* **58**, 677–689 (2015).
- 2 Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
- 3 Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
- 4 Bai, X.-C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
- 5 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
- 6 Pfab, J., Phan, N. M. & Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl Acad. Sci. USA* **118**, e2017525118 (2021).
- 7 Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).

- 8 Su, B., Huang, K., Peng, Z., Amunts, A. & Yang, J. CryoAtom improves model building for cryo-EM. *Nat. Struct. Mol. Biol.* **33**, 351–361 (2026).
- 9 Xu, Y. & Dang, S. Recent technical advances in sample preparation for single-particle cryo-EM. *Front. Mol. Biosci.* **9**, 892459 (2022).
- 10 Doerr, A. Cryo-electron tomography. *Nat. Methods* **14**, 34–34 (2017).
- 11 Nogales, E. & Mahamid, J. Bridging structural and cell biology with cryo-electron microscopy. *Nature* **628**, 47–56 (2024).
- 12 Asano, S., Engel, B. D. & Baumeister, W. In situ cryo-electron tomography: a post-reductionist approach to structural biology. *J. Mol. Biol.* **428**, 332–343 (2016).
- 13 Young, L. N. & Villa, E. Bringing structure to cell biology with cryo-electron tomography. *Annu. Rev. Biophys.* **52**, 573–595 (2023).
- 14 Wang, T. *et al.* CryoSeek: A strategy for bioentity discovery using cryoelectron microscopy. *Proc. Natl Acad. Sci. USA* **121**, e2417046121 (2024).
- 15 Wang, T. *et al.* CryoSeek II: Cryo-EM analysis of glycofibrils from freshwater reveals well-structured glycans coating linear tetrapeptide repeats. *Proc. Natl Acad. Sci. USA* **122**, e2423943122 (2025).
- 16 Ho, C.-M. *et al.* Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods* **17**, 79–85 (2020).
- 17 Gao, J. *et al.* DomainFit: Identification of protein domains in cryo-EM maps at intermediate resolution using AlphaFold2-predicted models. *Structure* **32**, 1248–1259.e5 (2024).
- 18 Taigman, Y., Yang, M., Ranzato, M. A. & Wolf, L. Deepface: Closing the gap to human-level performance in face verification. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1701–1708 (2014).
- 19 Schroff, F., Kalenichenko, D. & Philbin, J. Facenet: A unified embedding for face recognition and clustering. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 815–823 (2015).
- 20 Orengo, C. A. *et al.* CATH—a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1109 (1997).
- 21 Greene, L. H. *et al.* The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res.* **35**, D291–D297 (2007).

- 22 Fox, N. K., Brenner, S. E. & Chandonia, J.-M. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.* **42**, D304–D309 (2014).
- 23 Lawson, C. L. *et al.* EMDDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403 (2016).
- 24 Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, 234–241 (2015).
- 25 He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 770–778 (2016).
- 26 Zhang, K. *et al.* Practical blind image denoising via Swin-Conv-UNet and data synthesis. *Mach. Intell. Res.* **20**, 822–836 (2023).
- 27 Ji, S., Xu, W., Yang, M. & Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 221–231 (2012).
- 28 Lin, M., Chen, Q. & Yan, S. Network in network. *arXiv*, arXiv:1312.4400 (2013).
- 29 Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. *arXiv*, arXiv:1607.06450 (2016).
- 30 Vaswani, A. *et al.* Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30** (2017).
- 31 Douze, M. *et al.* The Faiss Library. *IEEE Trans. Big Data* **12**, 346–361 (2025).
- 32 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- 33 Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).
- 34 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
- 35 Taheri-Ledari, M., Zandieh, A., Shariatpanahi, S. P. & Eslahchi, C. Assignment of structural domains in proteins using diffusion kernels on graphs. *BMC Bioinformatics* **23**, 369 (2022).
- 36 Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026 – 1028 (2017).
- 37 Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018).

- 38 Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995–1014 (2014).
- 39 Liu, X. *et al.* Self-supervised learning: Generative or contrastive. *IEEE Trans. Knowl. Data Eng.* **35**, 857–876 (2021).
- 40 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- 41 Radford, A. *et al.* Learning transferable visual models from natural language supervision. *Proc. 38th Int. Conf. Mach. Learn.*, 8748–8763 (2021).
- 42 Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *Proc. 36th Int. Conf. Mach. Learn.*, 6105–6114 (2019).
- 43 Tsipras, D., Santurkar, S., Engstrom, L., Ilyas, A. & Madry, A. From imagenet to image classification: Contextualizing progress on benchmarks. *Proc. 37th Int. Conf. Mach. Learn.*, 9625–9635 (2020).
- 44 Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 (2016).
- 45 Iudin, A. *et al.* EMPIAR: the electron microscopy public image archive. *Nucleic Acids Res.* **51**, D1503–D1511 (2023).
- 46 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- 47 Manning, C. D. Introduction to Information Retrieval. *Cambridge Univ. Press* (2008).
- 48 Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 (2015).
- 49 Briggs, J. A. Structural biology in situ—the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**, 261–267 (2013).
- 50 Wan, W. & Briggs, J. A. Cryo-electron tomography and subtomogram averaging. *Methods Enzymol.* **579**, 329–367 (2016).
- 51 Weaver, J. *et al.* GroEL actively stimulates folding of the endogenous substrate protein PepQ. *Nat. Commun.* **8**, 15934 (2017).

- 52 Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586 (1994).
- 53 Meena, S. R. & Saxena, A. K. Crystal structure of apo-GroEL structure. PDB-4HEL (2013). <https://doi.org/4hel>
- 54 Ni, T. *et al.* Structure of native HIV-1 cores and their interactions with IP6 and CypA. *Sci. Adv.* **7**, eabj5715 (2021).
- 55 Dos Santos, N. F. *et al.* Lenacapavir allosterically remodels the HIV-1 capsid. *bioRxiv*, 2026.01.05.697065 (2026).
- 56 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
- 57 Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).
- 58 Zhang, C., Freddolino, L. & Zhang, Y. A graphic and command line protocol for quick and accurate comparisons of protein and nucleic acid structures with US-align. *Nat. Protoc.* **21**, 517–541 (2026).
- 59 Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score=0.5? *Bioinformatics* **26**, 889–895 (2010).
- 60 Orengo, C. A., Jones, D. T. & Thornton, J. M. Protein superfamilies and domain superfolds. *Nature* **372**, 631–634 (1994).
- 61 Holm, L. & Sander, C. Mapping the protein universe. *Science* **273**, 595–602 (1996).
- 62 Li, F. *et al.* Structural atlas of Pakpunavirus P7-1 reveals determinants of virion stability and genome ejection. *Commun. Biol.* **9**, 913 (2026).
- 63 Subramanian, S., Kerns, H. R., Braverman, S. G. & Doore, S. M. The structure of Shigella virus Sf14 reveals the presence of two decoration proteins and two long tail fibers. *Commun. Biol.* **8**, 222 (2025).
- 64 Hay, I. M. *et al.* Structural basis for a phosphoinositide-driven mTORC2-AKT positive feedback loop. *bioRxiv*, 2026.01.08.698367 (2026).
- 65 Ll  cer, J. L. *et al.* Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* **59**, 399–412 (2015).
- 66 Harris, J. A. *et al.* Selective G protein signaling driven by substance P–neurokinin receptor dynamics. *Nat. Chem. Biol.* **18**, 109–115 (2022).

67 Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).

Figure legends

Fig1

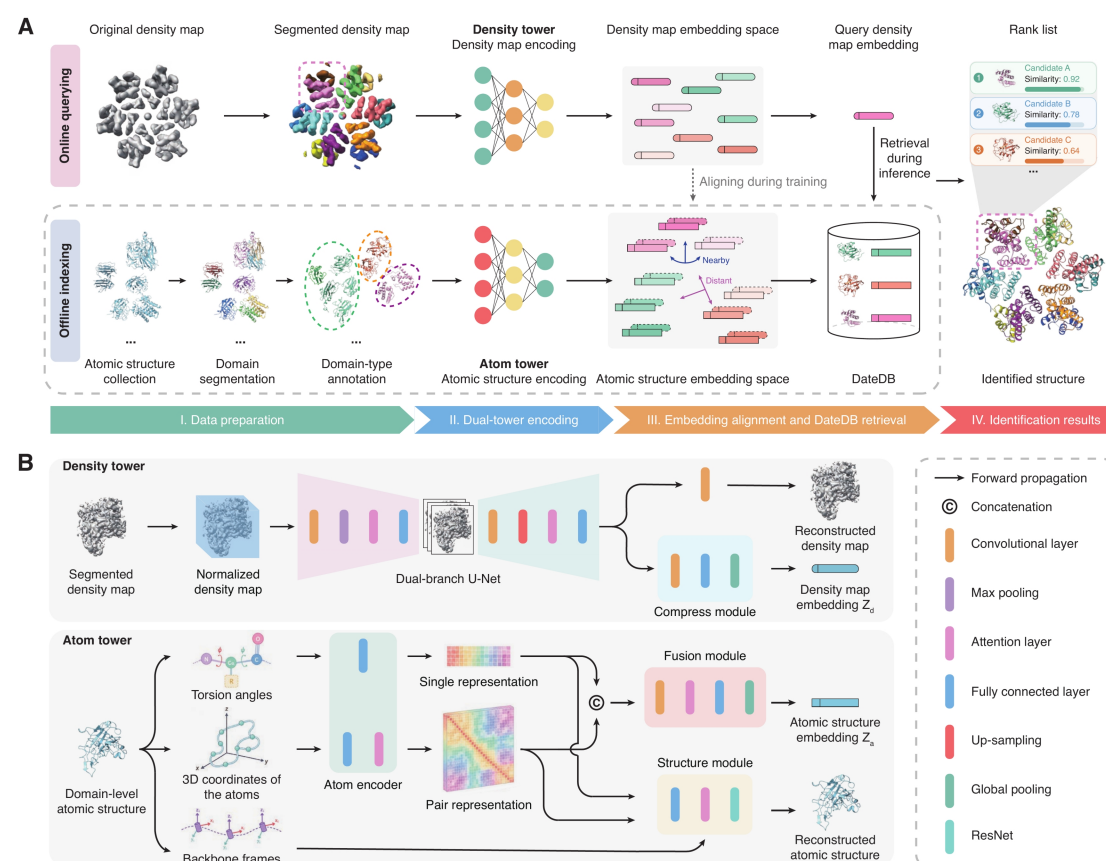


Figure 1. Overview of CryoNet.Discovery.

A. Workflow of CryoNet.Discovery. CryoNet.Discovery starts from an experimental 3D cryo-EM density map, which is segmented into domain density maps. Each segmented map is encoded into a density map embedding, which is used to query against a pre-indexed reference gallery (DateDB) to retrieve the corresponding protein domain type and atomic structure. The gray dashed box illustrates the construction of DateDB, which is built from curated collections of domain atomic structures (e.g., CATH or SCOPe); each domain structure is encoded and stored to form the reference gallery (DateDB) for subsequent retrieval.

B. Architecture of CryoNet.Discovery. The density tower takes a density map as input, and extracts both local and global features through a dual-branch U-Net architecture. It outputs a density map embedding (Z_d) along with a reconstructed density map through the compress module and convolutional layer. The atom tower takes an atomic structure as input, extracts side-chain torsion angles, atomic coordinates, and backbone frames, and generates an atomic structure embedding (Z_a) together with a reconstructed atomic model through the atom encoder, structure module, and fusion module.

Fig2

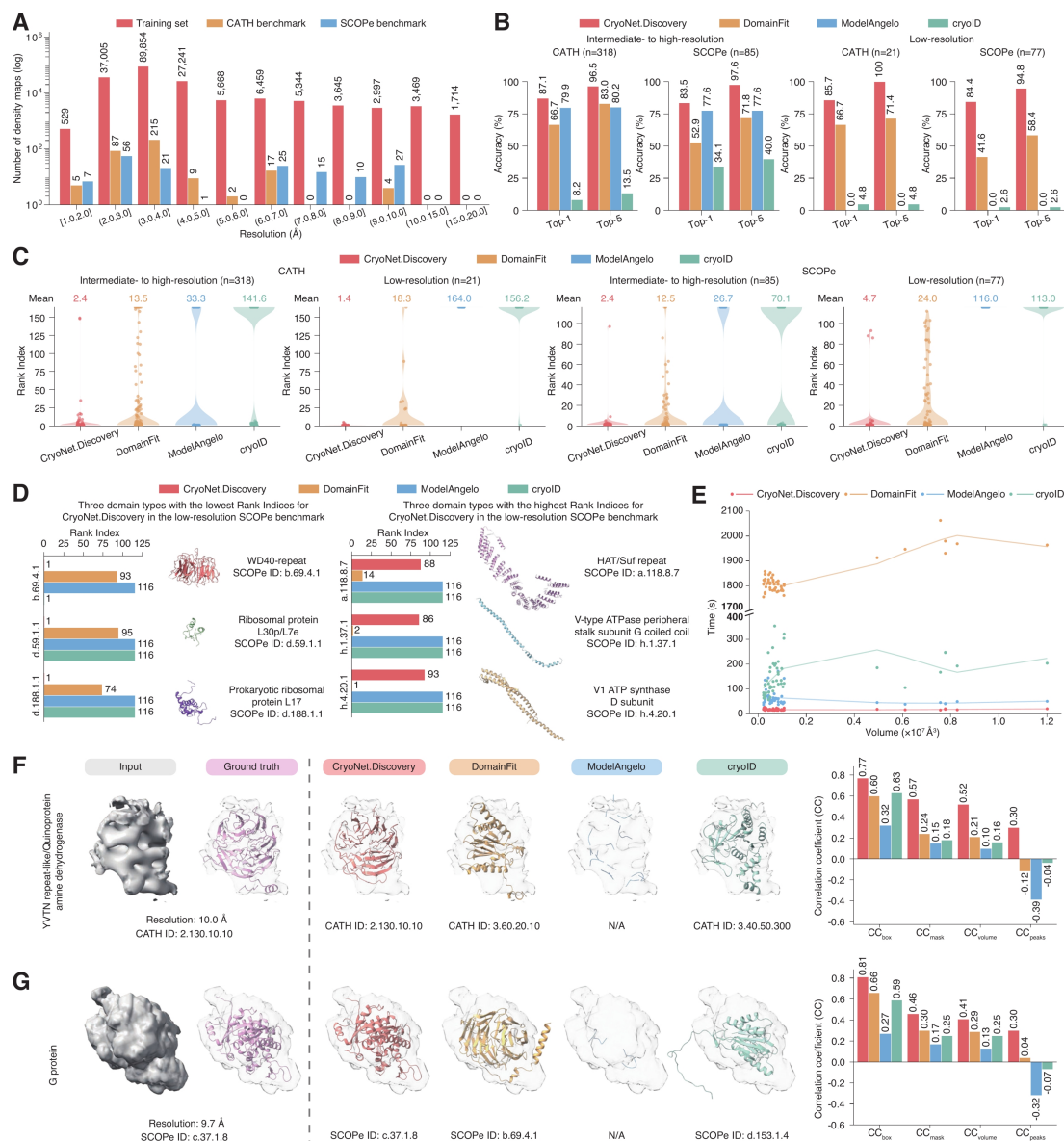


Figure 2. Performance of CryoNet.Discovery.

A. Resolution distribution of the training set and benchmarks. Red bars indicate density maps used during training, orange bars indicate density maps in CATH benchmark, and blue bars indicate density maps in SCOPe benchmark.

B. Overall performance comparison on the CATH and SCOPe benchmarks. CryoNet.Discovery is compared with DomainFit, ModelAngelo, and cryoID. Top-1

and Top-5 accuracies are reported, with the number of benchmark queries indicated above each panel.

C. Distribution of Rank Indices for structure identification on the CATH and SCOPe benchmarks. Each point represents one density map, and the violin plots summarize the distribution. The mean Rank Index is shown above each method; lower values indicate better ranking of the correct domain type.

D. Three domain types with the lowest/highest Rank Indices for CryoNet.Discovery in the low-resolution SCOPe benchmark. Bars show the Rank Index achieved by each method for each domain type, with representative atomic structures shown to the right of each bar plot..

E. Comparison of inference efficiency. The relationship between inference time and input density map voxel volume is shown for CryoNet.Discovery, DomainFit, ModelAngelo, and cryoID. The x-axis represents voxel volume, and the y-axis denotes total runtime from input to output, including pre-processing, inference, and post-processing. There are 50 samples, and each point corresponds to one sample, with solid lines indicating overall trends.

F–G. Examples of structure identification on low-resolution density regions. The input density, reference structure, and Top-1 results of CryoNet.Discovery, DomainFit, ModelAngelo, and cryoID are shown for a CATH-annotated YWTD repeat-like/quinoprotein amine dehydrogenase domain (F) and a SCOPe-annotated G-protein domain (G), with map–model agreement quantified using CC_{box} , CC_{mask} , CC_{volume} , and CC_{peaks} . For ModelAngelo, the model it built is retained when sequence assignment failed, with the domain type labeled N/A.

Fig3

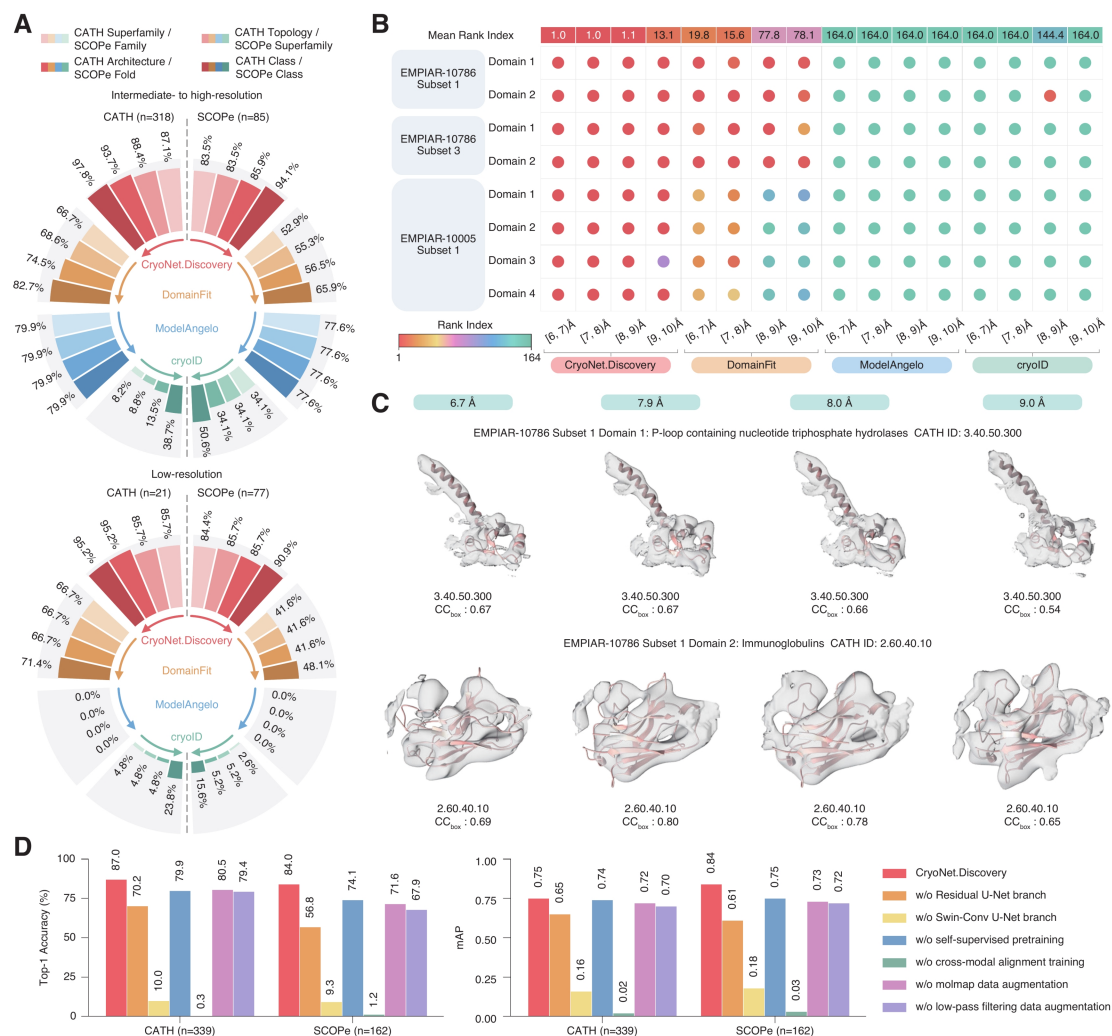


Figure 3. Performance at different classification levels, resolution, and ablation studies of CryoNet.Discovery.

A. Structure identification performance across hierarchical classification levels for CATH and SCOPe benchmarks. Radial bars show Top-1 accuracy across progressively coarser classification levels, including CATH homologous superfamily, topology, architecture, and class, and SCOPe family, superfamily, fold, and class.

B. Robustness of structure identification to resolution variations. Three EMPIAR particle-image sets are reconstructed using cryoSPARC at different down-sampling levels, yielding density maps across four resolution ranges ([6,7) Å, [7,8) Å, [8,9) Å, and [9,10) Å). Eight groups of CATH-annotated domain regions are extracted and

identified using CryoNet.Discovery, DomainFit, ModelAngelo, and cryoID. Dot colors encode the Rank Index of the correct CATH domain type from 1 to 164, with lower values representing better rankings.

C. Visual examples of robustness to resolution variations for two groups of CATH-annotated domain regions from EMPIAR-10786 Subset 1. Segmented density regions and the fitted Top-1 structures identified by CryoNet.Discovery are shown across 6.7 Å, 7.9 Å, 8.0 Å, and 9.0 Å reconstructions, together with the corresponding domain type and CC_{box} values.

D. Ablation study of CryoNet.Discovery on CATH and SCOPe benchmarks. Top-1 accuracy and mean average precision (mAP) are compared between the full model and variants removing the Residual U-Net branch, the Swin-Conv U-Net branch, self-supervised pretraining, cross-modality alignment training, molmap data augmentation, or low-pass filtering data augmentation.

Fig4

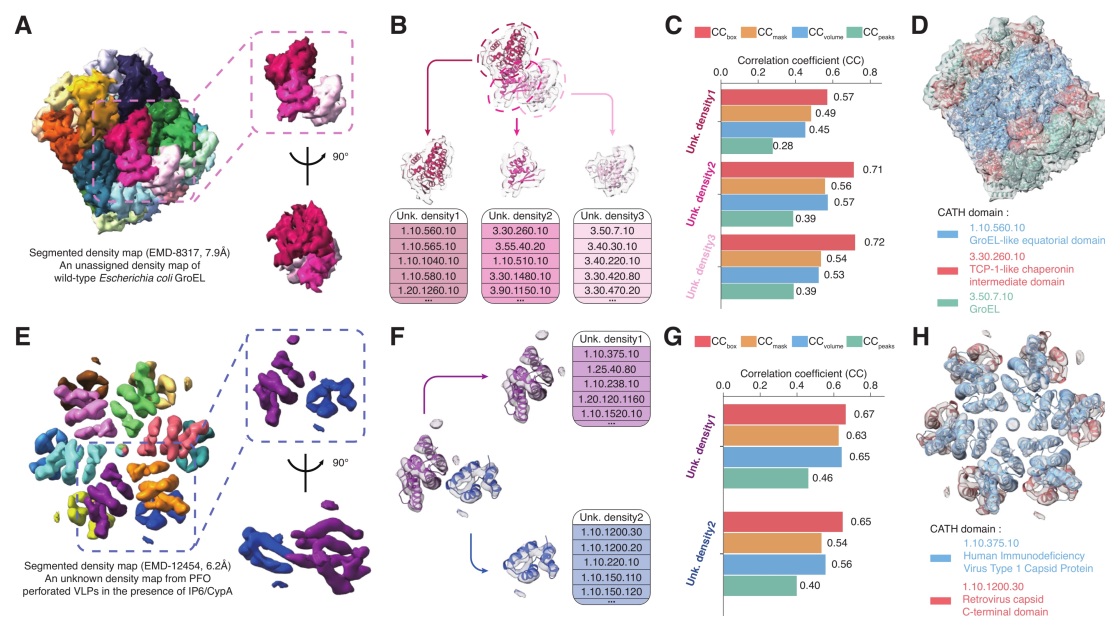


Figure 4. Identification from unassigned single-particle cryo-EM and cryo-ET density maps.

A. Density-region segmentation of an unassigned SPA density map. The original SPA density map is an unassigned 7.9 Å cryo-EM map of wild-type *Escherichia coli* GroEL deposited in the EMDB (EMD-8317). No atomic model was provided for this density map in the original study. The original density map was segmented into individual density regions, which are shown in distinct colors. Representative density regions from the symmetric assembly are highlighted and displayed from different orientations.

B. Structure identification for the three representative segmented density regions using CryoNet.Discovery. For each selected density region, the Top-5 candidate domain type IDs returned from the DateDB-CATH-full are listed, and the Top-1 identified atomic structure is shown after fitting to the density. Dashed outlines mark the density regions selected from the segmented map for identification.

C. Quantitative assessment of density-model agreement for the three representative density regions. Bar charts report four map-model correlation

coefficient metrics computed between each density region and its fitted Top-1 identified atomic structure: CC_{box} (red), CC_{mask} (orange), CC_{volume} (blue), and CC_{peaks} (green). Bars are grouped by density region (Unk. density1–3), and all metrics range from 0 to 1.

D. Organization of the identified domain types for the whole original density map. All segmented density regions are independently identified using CryoNet.Discovery and the identified domain types are mapped back onto the whole original density map. Different colors indicate different domain types, with blue corresponding to the GroEL-like equatorial domain (CATH ID: 1.10.560.10) according to the CATH classification, red to the TCP-1-like chaperonin intermediate domain (CATH ID: 3.30.260.10), and green to the GroEL (CATH ID: 3.50.7.10).

E. Density-region segmentation of an unassigned subtomogram-averaging density map. The original experimental density map is an unassigned 6.2 Å subtomogram-averaging density map from PFO perforated VLPs in the presence of IP6/CypA, deposited in the EMDB (EMD-12454). No atomic model was provided for this density map in the original study. The map was segmented into individual density regions, which are shown in distinct colors. Representative density regions from the symmetric assembly are highlighted and displayed from different orientations.

F. Structure identification for the two representative density regions using CryoNet.Discovery. For each selected density region, the Top-5 candidate domain type IDs returned from the DateDB-CATH-full are listed, and the Top-1 identified atomic structure is shown after fitting to the density. Dashed outlines mark the density regions selected from the segmented map for identification.

G. Quantitative assessment of density–model agreement for the two representative density regions. Bar charts report four map-model correlation coefficient metrics computed between each density region and its fitted Top-1 identified atomic structure: CC_{box} (red), CC_{mask} (orange), CC_{volume} (blue), and CC_{peaks}

(green). Bars are grouped by density region (Unk. density1–2), and all metrics range from 0 to 1.

H. Organization of the identified domain types for the whole original density map. All segmented density regions are independently identified using CryoNet.Discovery and the identified domain types are mapped back onto the whole original density map. Different colors indicate different domain types, with blue corresponding to the Human Immunodeficiency Virus type 1 capsid protein (CATH ID: 1.10.375.10) according to the CATH classification and red to the retrovirus capsid C-terminal domain (CATH ID: 1.10.1200.30).

Fig5

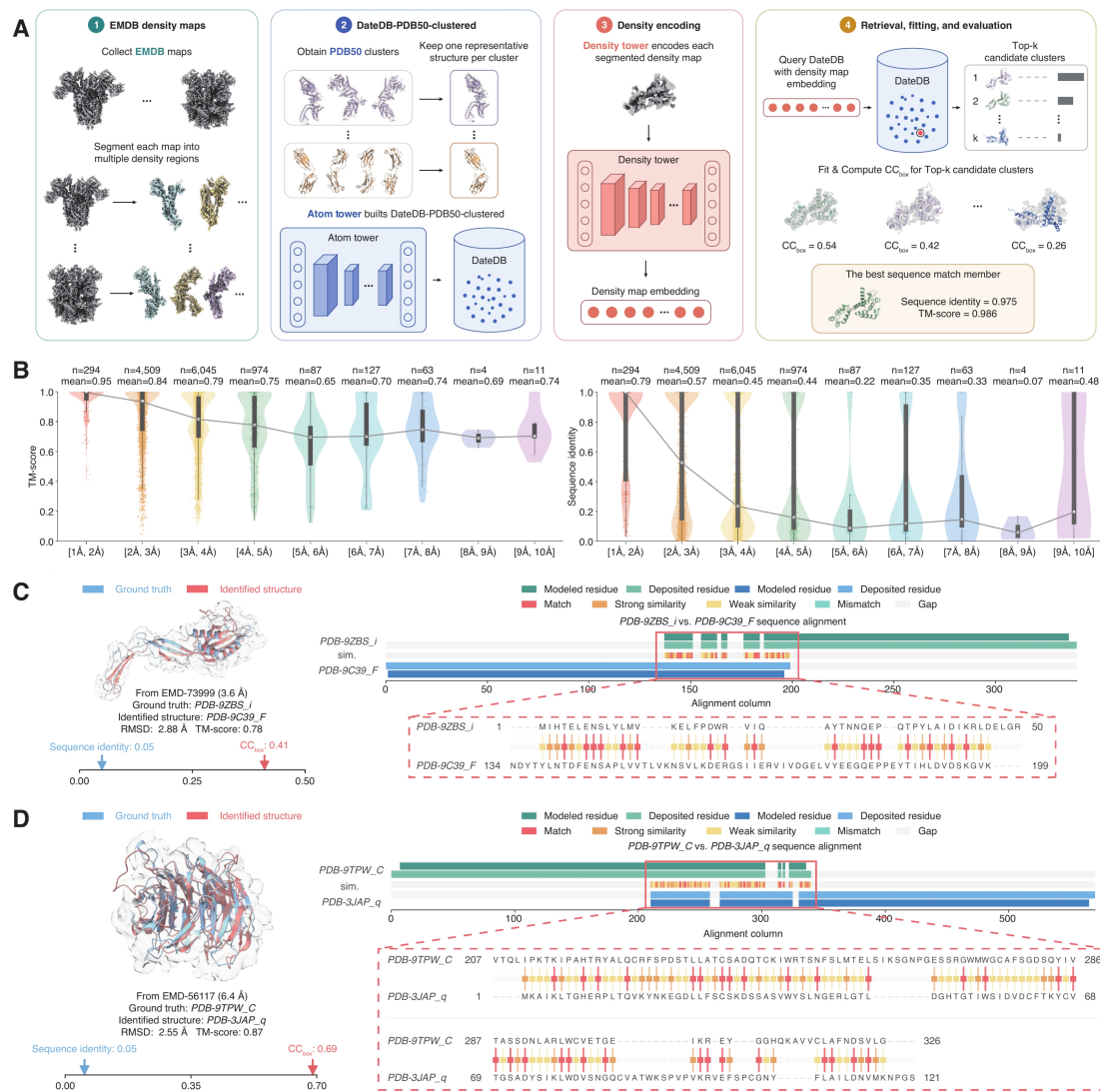


Figure 5. Large-scale whole-chain structure identification across the EMDB.

A. EMDB-scale structure-identification evaluation pipeline. EMDB density maps are collected and segmented into individual density regions. A PDB50-clustered single-chain DateDB (DateDB-PDB50-clustered) is constructed by retaining one representative structure from each PDB50 sequence cluster. These representative structures are encoded by the atom tower and indexed in DateDB. Each query density region is encoded by the density tower and searched against DateDB. The retrieved Top-k candidate clusters are reranked by density–model agreement (CC_{box}), expanded

to their member chains, and evaluated against the deposited ground-truth chain using sequence identity and TM-score.

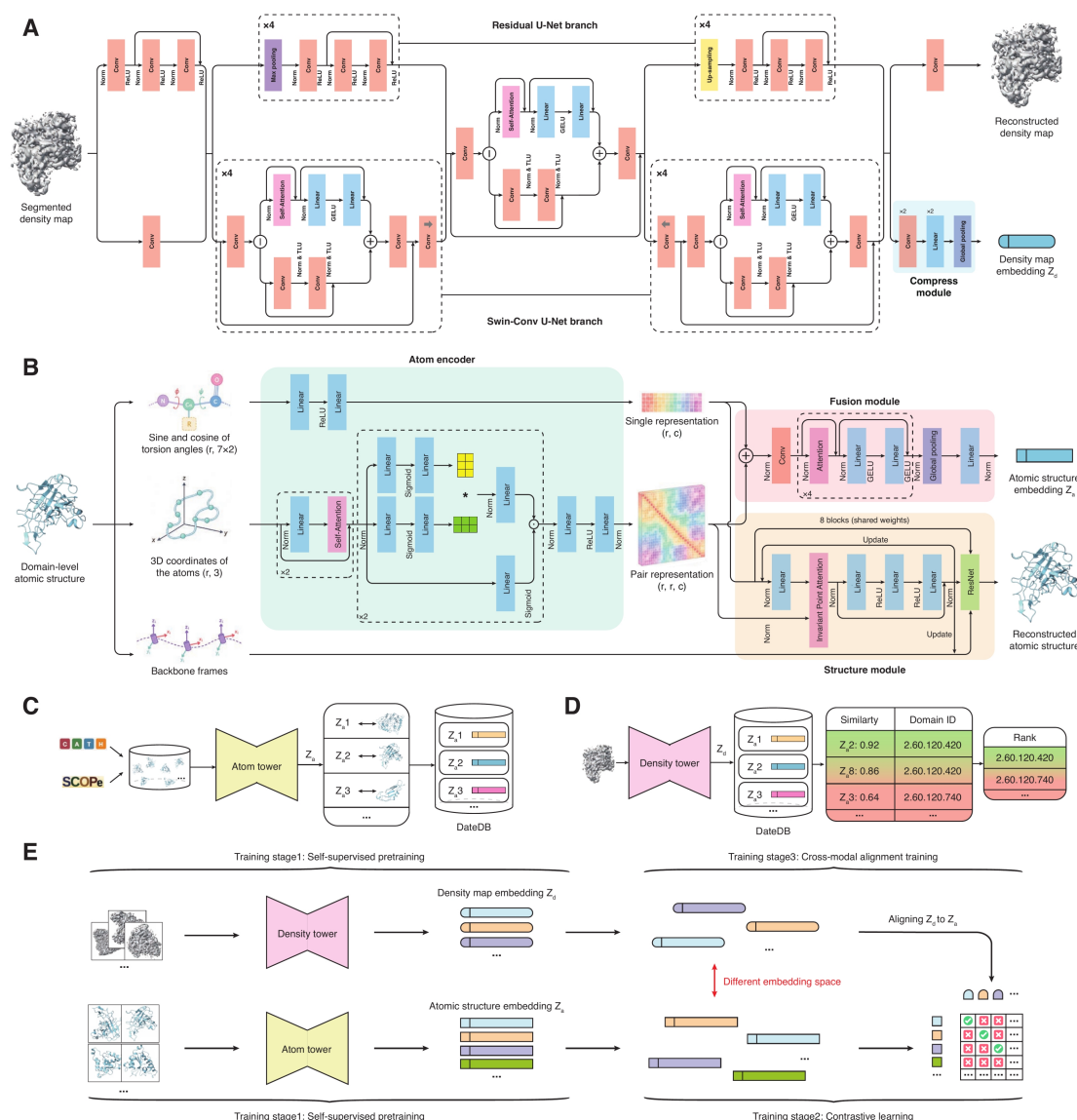
B. Resolution-stratified structure-identification analysis. CryoNet.Discovery is applied to 12,114 density regions from 1,488 cryo-EM density maps deposited in the EMDB on or after January 1, 2026. Each query is searched against DateDB-PDB50-clustered, followed by map-model reranking and cluster expansion. The left panel shows structure-alignment-based TM-scores between the deposited ground-truth chains and selected candidate chains across resolution bins, whereas the right panel shows the corresponding sequence identities calculated by EMBOSS Needle global alignments. Violin plots show the distributions, individual points represent query density regions, grey boxes indicate interquartile ranges, and grey lines connect bin medians. The number of queries and mean TM-score/sequence identity are shown above each resolution bin.

C. Sequence-divergent structural match in a phage neck scaffold. The deposited ground-truth chain *PDB-9ZBS_i* and the identified chain *PDB-9C39_F* are shown after fitting to the density region (a region from EMD-73999, 3.6 Å) in the left panel. RMSD, TM-score, CC_{box}, and sequence identity are indicated. The right panel shows the EMBOSS Needle global sequence alignment between *PDB-9ZBS_i* and *PDB-9C39_F*, with modeled and deposited residues. Matches, similarities, mismatches, and gaps are marked, and the dashed box marks the aligned region in detail.

D. Sequence-divergent structural match in a WD40/β-propeller-like scaffold. The deposited ground-truth chain *PDB-9TPW_C* and the identified chain *PDB-3JAP_q* are shown after fitting to the density region (a region from EMD-56117, 6.4 Å) in the left panel. RMSD, TM-score, CC_{box}, and sequence identity are indicated. The right panel shows the EMBOSS Needle global sequence alignment between *PDB-9TPW_C* and *PDB-3JAP_q*, with modeled and deposited residues. Matches, similarities,

mismatches, and gaps are marked, and the dashed box marks the aligned region in detail.

Supplementary Fig1



Supplementary Figure 1. Detailed architecture, training strategy, and workflow of CryoNet.Discovery

A. Density tower architecture. The density tower encodes density maps into density map embeddings. The network consists of a dual-branch U-Net (Residual U-Net branch and Swin-Conv U-Net branch) followed by a convolution layer to reconstruct the input density map and a compression module to generate a density map embedding Z_d .

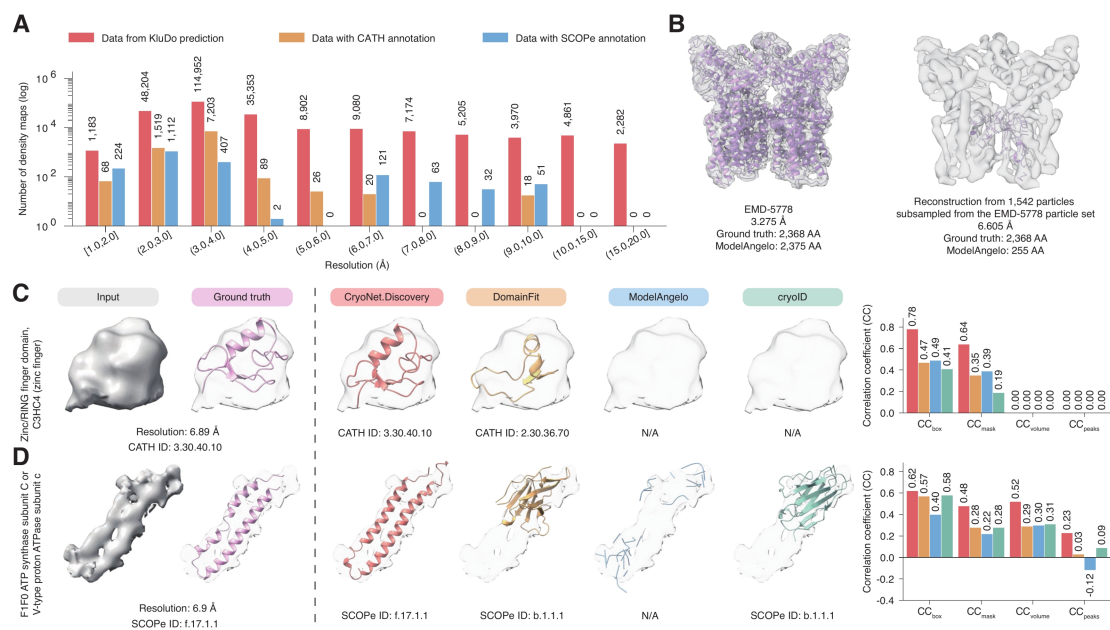
B. Atom tower architecture. The atom tower encodes atomic structures into atomic structure embeddings. Atomic structures are encoded into single and pair representations, which are then processed by a structure module to reconstruct the input atomic structure and a fusion module to generate an atomic structure embedding Z_a .

C. Construction of DateDB. Collected domain atomic structures are encoded by the atom tower and stored as embeddings to form the reference gallery used for retrieval.

D. Structure identification workflow of CryoNet.Discovery. Given a query density region, the density tower generates a density map embedding, which is compared against all atomic structure embeddings in DateDB. Candidate domain types and their corresponding atomic structures are returned according to the similarity ranking.

E. Training stages. CryoNet.Discovery is trained in three stages: self-supervised pretraining within each modality for the density tower and atom tower, contrastive learning on the atomic structure modality to enhance domain separability, and cross-modal alignment training to align density map embeddings (Z_d) with atomic structure embeddings (Z_a).

Supplementary Fig2



Supplementary Figure 2. Data distribution and more examples of structure identification

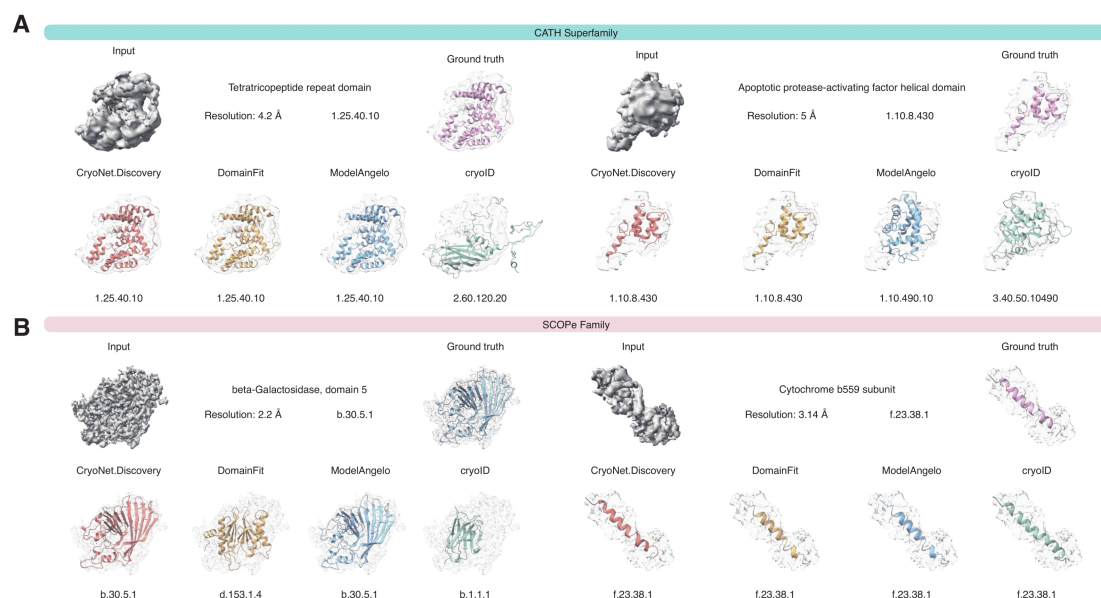
A. Resolution distribution of domain density maps. Red bars indicate density maps obtained from KluDo prediction, orange bars indicate density maps with CATH annotations, and blue bars indicate density maps with SCOPe annotations.

B. Comparison of ModelAngelo outputs at different resolutions. ModelAngelo-built models are shown for the deposited EMD-5778 map at 3.275 Å and for a 6.605 Å map reconstructed from the same particle set. Grey surfaces indicate density maps, and purple ribbons indicate ModelAngelo models. The residue counts of the deposited ground-truth structure and the ModelAngelo-built model are indicated below each map.

C–D. Examples of structure identification on low-resolution density regions. The input density, reference structure, and Top-1 results from CryoNet.Discovery, DomainFit, ModelAngelo, and cryoID are shown for a CATH-annotated Zinc/RING finger domain (C) and a SCOPe-annotated F1F0 ATP synthase subunit C or V-type

proton ATPase subunit c domain (D), with map–model agreement quantified using CC_{box} , CC_{mask} , CC_{volume} , and CC_{peaks} . For ModelAngelo, the model it built is retained when sequence assignment failed, with the domain type labeled N/A, whereas ModelAngelo model-building failures and missing cryoID outputs are shown as empty atomic structures.

Supplementary Fig3

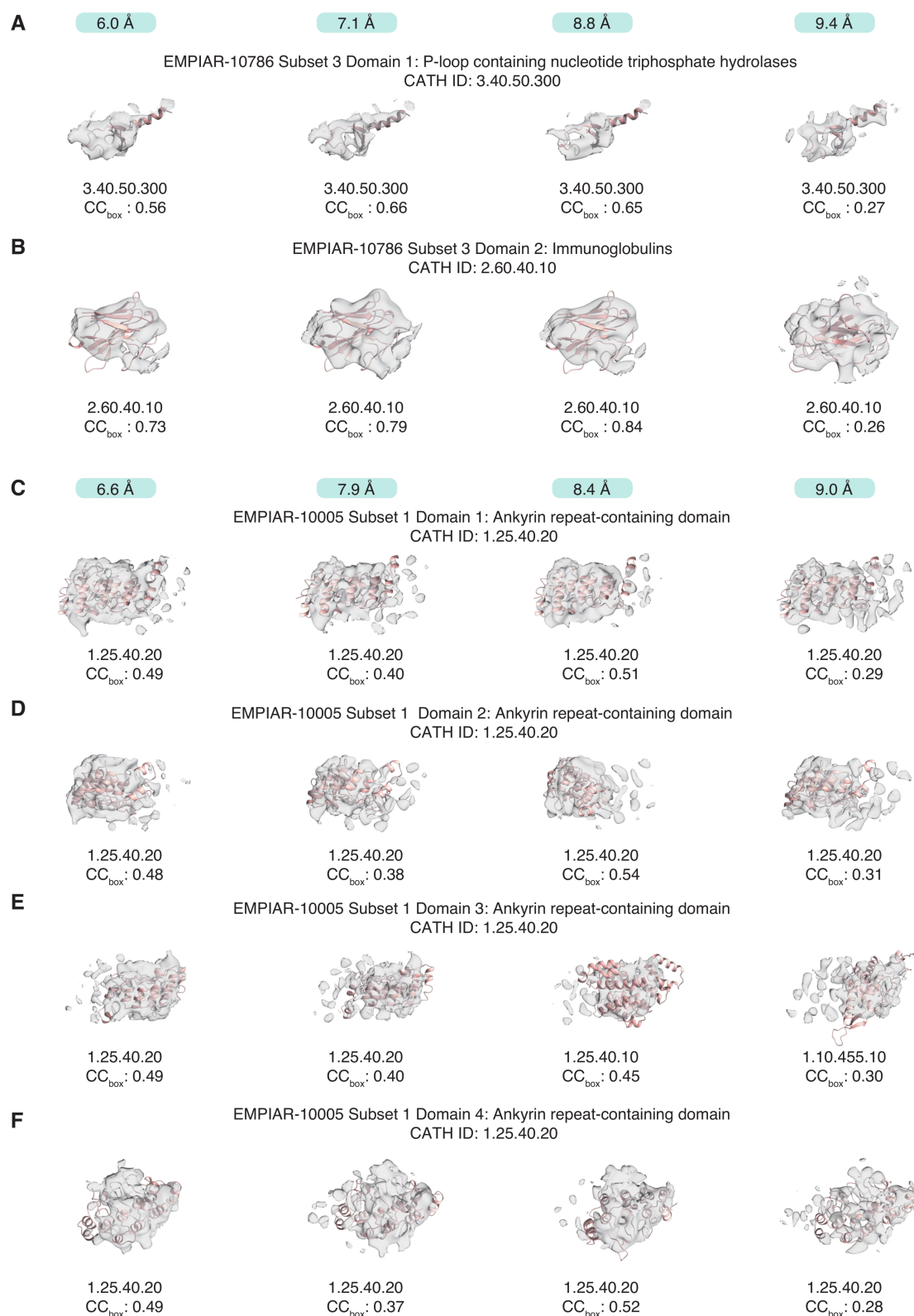


Supplementary Figure 3. Examples of structure identification on high-resolution density regions.

A-B. Two CATH-annotated cases (A) and two SCOPe-annotated cases (B) are shown. For each example, the top row shows the input density region, together with its global resolution and the corresponding reference atomic structure annotated with the domain type and domain type ID. The bottom row shows the Top-1 identified domain atomic structures by CryoNet.Discovery, DomainFit, ModelAngelo, and cryoID, respectively, together with their associated domain type IDs.

In panel A (CATH superfamilies), the examples include a tetratricopeptide repeat domain (4.2 Å; CATH ID: 1.25.40.10) and an apoptotic protease-activating factor helical domain (5.0 Å; CATH ID: 1.10.8.430). In panel B (SCOPe families), the examples include a beta-galactosidase, domain 5 (2.2 Å; SCOPe ID: b.30.5.1) and a cytochrome b559 subunit (3.14 Å; SCOPe ID: f.23.38.1).

Supplementary Fig4

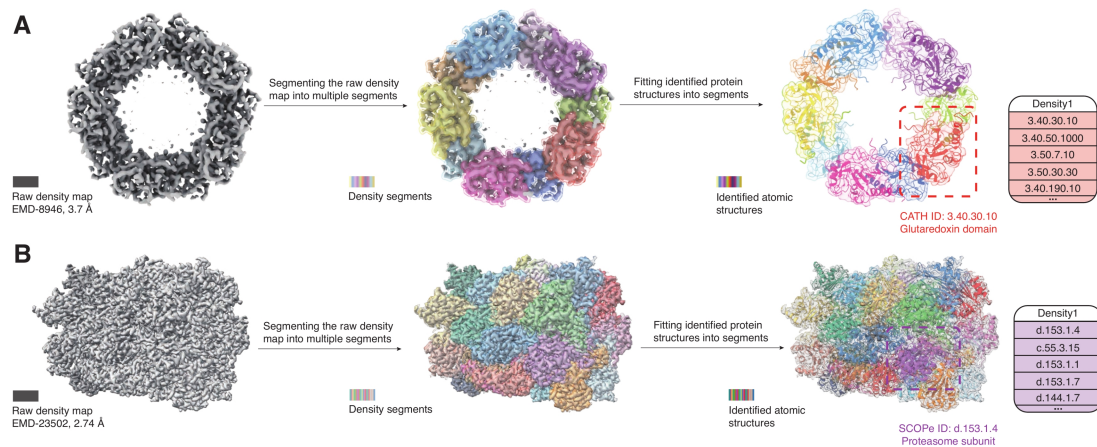


Supplementary Figure 4. Additional visual examples of robustness to resolution variations.

A–B. Segments from EMPIAR-10786 Subset 3 reconstructions. The two groups of segmented density regions correspond to a P-loop containing nucleotide triphosphate hydrolase domain (CATH ID: 3.40.50.300, A) and an immunoglobulin domain (CATH ID: 2.60.40.10, B). Each row shows the corresponding segmented regions obtained from the density maps at different global resolutions (6.0 Å, 7.1 Å, 8.8 Å, and 9.4 Å), together with the fitted Top-1 identified atomic structure and the corresponding domain type identified by CryoNet.Discovery, as well as the corresponding CC_{box}.

C–F. Segments from EMPIAR-10005 Subset 1 re-reconstructions. Each group of segmented density regions corresponds to an ankyrin repeat-containing domain annotated in CATH (CATH ID: 1.25.40.20). Each row shows the corresponding segmented regions obtained from the density maps at different global resolutions (6.6 Å, 7.9 Å, 8.4 Å, and 9.0 Å), together with the fitted Top-1 identified atomic structure and the corresponding domain type identified by CryoNet.Discovery, as well as the corresponding CC_{box}.

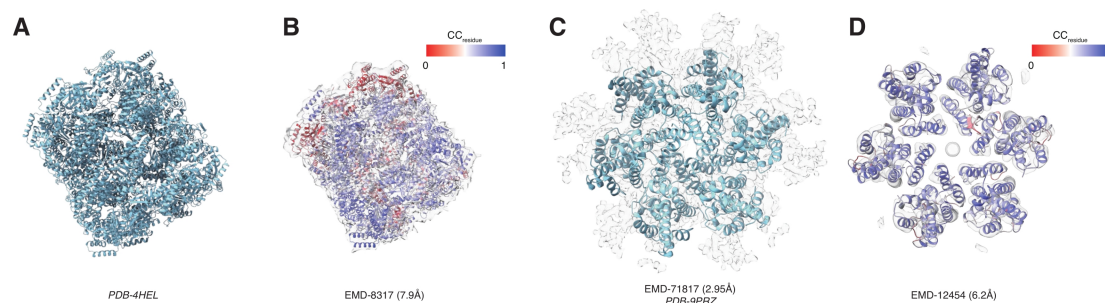
Supplementary Fig5



Supplementary Figure 5. Two known density maps with high-resolution.

A-B. For each panel, the left shows the original experimental density map (A, EMD-8946, 3.7 Å; B, EMD-23502, 2.74 Å). The middle shows the corresponding density regions obtained after segmentation, with different colors indicating distinct segmented regions. The right shows the identified domain atomic structures fitted back into the corresponding density regions. Dashed outlines highlight representative domains selected from the symmetric assemblies. For each highlighted density region, the Top-5 candidate domain type IDs returned from the DateDB-CATH-full (A) or DateDB-SCOPe-full (B) are listed. In panel A, the highlighted region corresponds to a glutaredoxin domain (CATH ID: 3.40.30.10), whereas in panel B, the highlighted region corresponds to a proteasome subunit (SCOPe ID: d.153.1.4).

Supplementary Fig6



Supplementary Figure 6. Supplementary validation of density-model fitting and domain assignment in Fig. 4.

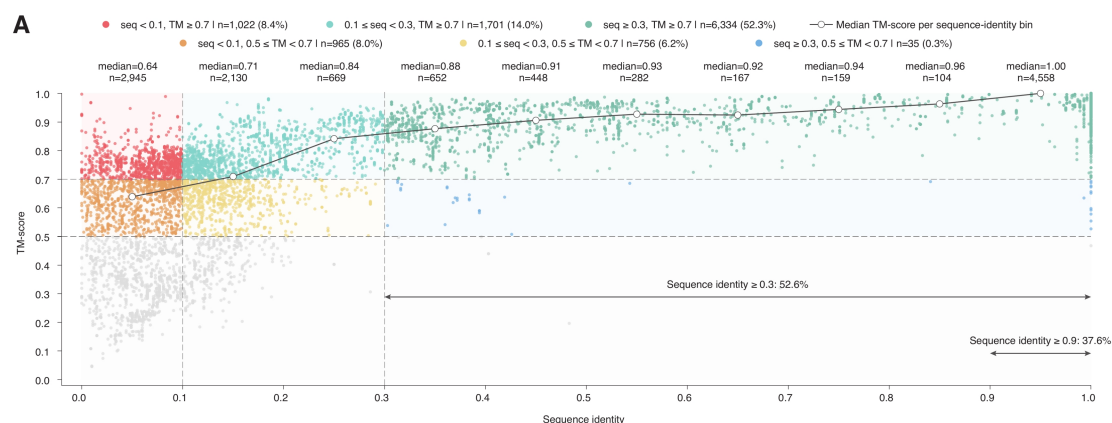
A. Startup model of EMD-8317. EMD-8317's startup model atomic model (*PDB-4HEL*) are shown.

B. Residue-level fitting consistency for EMD-8317. Atomic models fitted into the EMD-8317 density map in Fig. 4 are colored by residue-wise map-model correlation coefficient metric, CC_{residue} , from 0 to 1.

C. High-resolution reference for the capsid assignment. The EMD-71817 density map (2.95 Å) and its corresponding deposited atomic model (*PDB-9PRZ*) are shown.

D. Residue-level fitting consistency for EMD-12454. Atomic models fitted into the EMD-12454 density map in Fig. 4 are colored by CC_{residue} from 0 to 1.

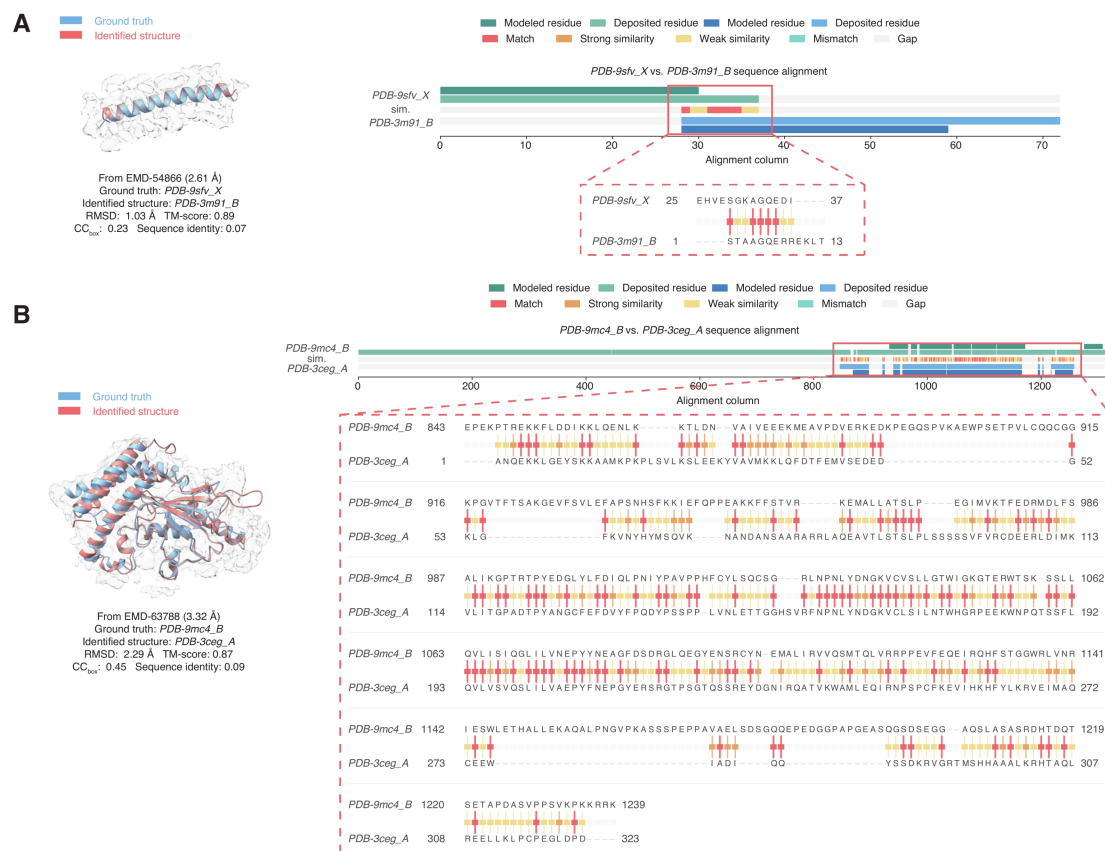
Supplementary Fig7



Supplementary Figure 7. Structure–sequence comparison of EMDB-scale whole-chain structure identification.

A. Relationship between sequence identity and structure-alignment-based TM-score. Each point represents one query density region and its selected candidate from the EMDB-scale structure-identification evaluation pipeline. Sequence identity is calculated by EMBOSS Needle global alignment, and TM-score is calculated by structure-based alignment. Dashed lines mark sequence-identity thresholds of 0.1 and 0.3 and TM-score thresholds of 0.5 and 0.7. The black line shows the median TM-score in each sequence-identity bin, with the median value and number of samples indicated above each bin.

Supplementary Fig8



Supplementary Figure 8. Representative failure modes among low-sequence-identity, high-TM-score identification cases.

A. Short-helical similarity case. For a density region from EMD-54866 (2.61 Å), the deposited ground-truth chain *PDB-9SFV_X* and the identified chain *PDB-3M91_B* are shown after fitting to the density region in the left panel. RMSD, TM-score, CC_{box}, and sequence identity are indicated. The right panel shows the EMBOS Needle global sequence alignment between *PDB-9SFV_X* and *PDB-3M91_B*, with modeled and deposited residues. Matches, similarities, mismatches, and gaps are marked, and the dashed box marks the aligned region in detail.

B. Sequence-coverage difference case. For a density region from EMD-63788 (3.32 Å), the deposited ground-truth chain *PDB-9MC4_B* and the identified chain *PDB-3CEG_A* are shown after fitting to the density region in the left panel. RMSD,

TM-score, CC_{box}, and sequence identity are indicated. The right panel shows the EMBOSS Needle global sequence alignment between *PDB-9MC4_B* and *PDB-3CEG_A*, with modeled and deposited residues. Matches, similarities, mismatches, and gaps are marked, and the dashed box marks the aligned region in detail.