©Copyright 2026
Chandramathi Murugadass

RNA Motif Detection, Structure Prediction and Analysis in Cryo-EM Density Maps: A Deep Learning Approach

Chandramathi Murugadass

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Computer Science and Software Engineering

University of Washington

2026

Committee:

Dong Si

Wooyoung Kim

Jie Hou

Program Authorized to Offer Degree:

Computing and Software Systems

University of Washington

Abstract

RNA Motif Detection, Structure Prediction and Analysis in Cryo-EM Density Maps: A Deep Learning Approach

Chandramathi Murugadass

Chair of the Supervisory Committee:
Associate Professor Dong Si
Computing and Software Systems

RNA function is closely linked to recurring secondary structural motifs, yet automated RNA three-dimensional structure determination from cryo-EM density remains limited by the lack of large, motif-resolved training datasets.

This thesis develops a motif-aware pipeline for RNA structure prediction from cryo-EM density maps through three contributions. First, it introduces 3D cryo-EM RNA Motif, an open-source dataset of over 100,000 motif-resolved density segments spanning 25 RNA secondary-structure classes and 1.5–34.0 Å resolutions, with voxel-level backbone, ribose, and nucleobase labels validated using cross-correlation and Q-score measures. Second, it presents Motif 3D CNN, which classifies five consolidated motif classes with 96.8% peak validation accuracy and 0.948 specificity on high-resolution maps. Third, it develops an end-to-end pipeline that localises motifs using FFT template matching, classifies candidates with Motif 3D CNN, segments the phosphate–sugar backbone using a MONAI UNETR transformer, and derives phosphorus-atom coordinates through skeletonisation and chain ordering. The resulting backbone probability maps achieve correlations of 0.1219–0.996 across hairpin and bulge motifs at varying resolutions. Together, these contributions provide the datasets, models, and pipeline required for motif-driven RNA modelling from cryo-EM density.

TABLE OF CONTENTS

Page

List of Figures iv

List of Tables v

Glossary vii

Chapter 1: Introduction 1
1.1 Motivation 2
1.2 Problem Statement 4
1.3 Contributions 4
1.4 Thesis Organisation 5

Chapter 2: Background and Related Work 6
2.1 RNA Structure and Motifs 6
2.1.1 Hairpins 6
2.1.2 Internal Loops 7
2.1.3 Bulges 9
2.2 Cryo-Electron Microscopy Density Maps 10
2.3 Related Work 11
2.3.1 RNA Motif Resources and Sequence-Based Motif Analysis 11
2.3.2 Deep Learning for cryo-EM Structure Modelling 13

Chapter 3: The 3DEM-RNA-Motif Dataset 15
3.1 Overview 15
3.2 Data Sources 15
3.3 Motif Classes 16
3.3.1 Retrieval 16

3.3.2 Segmentation . . . 16
3.3.3 Labelling . . . 17
3.3.4 Density Normalisation . . . 18
3.4 Quality Validation . . . 18
3.4.1 Cross-Correlation Fitting Score . . . 19
3.4.2 Q-Score Resolvability . . . 20
3.5 Dataset Statistics and Splits . . . 20

Chapter 4: Motif Classification . . . 25
4.1 Overview . . . 25
4.2 Consolidated Class Schema . . . 25
4.3 Motif 3D CNN Architecture . . . 27
4.3.1 Classifier Evaluation Metrics . . . 28
4.3.2 Confusion Matrix . . . 30

Chapter 5: Motif Localisation and Backbone Prediction in 3D Density Maps . . . 31
5.1 System Architecture . . . 31
5.2 Template Construction . . . 32
5.3 FFT Template Matching . . . 32
5.4 Template Matching Evaluation . . . 34
5.5 RNA Density Quality Check . . . 35
5.6 Motif Classification in Context . . . 35
5.7 UNETR Backbone Segmentation . . . 36
5.7.1 Architecture . . . 36
5.7.2 Training . . . 36
5.8 Post-processing and Backbone Tracing . . . 37

Chapter 6: Experimental Results . . . 39
6.1 Dataset Validation Results . . . 39
6.1.1 Cross-Correlation Scores . . . 39
6.1.2 Resolution-Stratified CC_{mask} Validation . . . 39
6.1.3 Representative Labelled Segments . . . 42
6.2 Motif Classification Results . . . 43
6.2.1 Training Convergence . . . 43

6.2.2 Blind-Test Classification Benchmark 44
6.2.3 Confusion Matrix Analysis 46
6.2.4 Prediction Results and Analysis 48
6.3 Inference Pipeline Results 48
6.3.1 Baseline Comparison 49
6.3.2 Template Matching and Localisation Performance 50
6.3.3 Motif Classification on Detected Candidates 50
6.3.4 Backbone Trace Quality 51
6.4 Discussion 51

Chapter 7: Conclusion and Future Work 55

Bibliography 59

Appendix A: Measure of fitness between cryo-EM map and atomic model of the curated dataset 64

Appendix B: Map–to–model FSC analysis 72

LIST OF FIGURES

Figure Number Page

1.1 Four schematic RNA secondary-structure motifs . . . 3

2.1 **Hairpins motif samples** . . . 7
2.2 **Asymmetric Loop** . . . 8
2.3 **Asymmetric Loop** . . . 9
2.4 **Bulge** . . . 10

3.1 **Motif segmentation and labeling** . . . 22
3.2 **Labelling a segmented motif** . . . 23
3.3 **Labeled RNA segment** . . . 24

4.1 **Motif-resolved RNA cryo-EM density maps** . . . 26
4.2 **3D CNN Classifier Workflow** . . . 29

6.1 **Average cross correlation** . . . 40
6.2 **Average Q-scores** . . . 42
6.3 **Model-map fourier shell correlation (FSC)** . . . 43
6.4 **Outlier motifs** . . . 44
6.5 Confusion matrix . . . 47
6.6 Predicted density map correlation results . . . 49
6.7 Traced Backbone . . . 52
6.8 Backbone Predicted Density . . . 53

B.1 High-resolution map–to–model FSC . . . 73
B.2 Mid-resolution map–to–model FSC . . . 74
B.3 Low-resolution map–to–model FSC . . . 75

LIST OF TABLES

Table Number Page

2.1 Comparison of representative RNA motif and sequence resources with respect to structural annotation and motif-level cryo-EM density availability (1). . . 12

3.1 Training, validation, and blind-test sample counts for the baseline 3D CNN motif classifier. Five models were trained using fold-specific training/validation splits and then evaluated on the blind test set. 21

4.1 Consolidated motif classes used for training Motif 3D CNN. The Subtypes column lists all 25 fine-grained classes that make up each consolidated class. 27

4.2 Confusion matrix for binary classification. 30

5.1 End-to-end pipeline stages for motif localisation and backbone prediction. . . 31

5.2 Software and hardware configuration for the inference pipeline. 32

5.3 UNETR model and training configuration. 37

6.1 Resolution-stratified CC_{mask} validation summary. n_{CC} is the number of motif samples in the CC_{mask} subset for each resolution bin; statistics summarise valid fit values within local motif-centred masks. 41

6.2 Key validation accuracy milestones during Motif 3D CNN training. 45

6.3 Per-class and macro-averaged blind-test performance of the Motif 3D CNN classifier, reported as mean ± standard deviation across the five trained models, each evaluated on a blind test set of 438 motif-density patches. 45

A.1 Distribution of segmented data (total counts of atomic model and density map) in the dataset, by motif class and nucleotide category. 66

A.2 Distribution of CC_{mask} validation samples by motif class and map resolution 67

A.3 Average cross correlation trends across motifs and resolution 68

A.4 Average Q-score correlation trend per motif and resolution 68

A.5 Blind-test confusion matrix for the 3D CNN model trained in Fold 3. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.826, Macro Specificity = 0.955. 69

A.6 Blind-test confusion matrix for the 3D CNN model trained in Fold 4. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.828, Macro Specificity = 0.956. 70

A.7 Blind-test confusion matrix for the 3D CNN model trained in Fold 5. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.856, Macro Specificity = 0.963. 70

A.8 Cryo-EM maps and fitted atomic models from which the RNA motifs were extracted. 71

B.1 Map–to–model FSC curves indexed by map resolution regime and randomly selected RNA motif type. Each cell gives the random motif instance under the motif group and a reference to its FSC plot. 72

GLOSSARY

cryo-EM: Cryogenic electron microscopy. An experimental technique for determining the three-dimensional structure of macromolecules by imaging frozen-hydrated samples with an electron beam.

MRC: Medical Research Council file format. A binary format for storing volumetric data such as electron density maps.

PDB: Protein Data Bank. An international repository of three-dimensional structural data for biological macromolecules.

EMDB: Electron Microscopy Data Bank. A public repository of three-dimensional cryo-EM density maps.

RNA COSSMOS: RNA Context of Secondary Structure Motifs. A database curated by the Znosko Lab that catalogs secondary structural motifs from RNA entries in the PDB.

MOTIF: A recurring three-dimensional structural element in an RNA molecule, including hairpin loops, internal loops, bulges, and junctions.

VOXEL: A volume element - the three-dimensional analogue of a pixel - used to represent density values in a discretised cryo-EM map.

FFT: Fast Fourier Transform. An algorithm for efficiently computing the discrete Fourier transform, used here for template cross-correlation.

CNN: Convolutional Neural Network. A class of deep learning model that applies learnable filters to structured grid data.

UNETR: U-Net Transformer. A segmentation architecture combining a Vision Transformer encoder with a CNN decoder connected via skip connections.

VIT: Vision Transformer. An architecture that processes image patches as sequences using multi-head self-attention.

Q-SCORE: Resolvability score that measures the similarity of map values around each atom relative to a Gaussian function for a well-resolved atom.

CC: Cross-correlation coefficient. A measure of similarity between an experimental density map and a model-derived density map on the same grid.

NMS: Non-maximum suppression. A post-processing step that removes duplicate detections from nearby positions in a score map.

DICECELOSS: A composite loss function combining Dice loss and cross-entropy loss used for voxel-level segmentation training.

MONAI: Medical Open Network for AI. A PyTorch-based open-source framework for deep learning in medical imaging.

DEEPTRACER: An existing deep learning pipeline for de-novo protein structure modelling from cryo-EM maps that is the foundation this work extends towards RNA.

CRYOREAD: A deep neural network method for de-novo nucleic acid backbone identification in cryo-EM maps.

ACKNOWLEDGMENTS

I am deeply grateful to my thesis committee chair, Associate Professor Dr. Dong Si, for his unwavering guidance, intellectual generosity, and patience throughout every phase of this research. His vision for motif-driven RNA structure prediction provided the foundation upon which this entire body of work rests.

I am equally thankful to Dr. Jie Hou for his invaluable scientific insights and continuous encouragement, and to the entire Data Analysis and Intelligent Systems (DAIS) research group at the University of Washington Bothell for daily discussions that sharpened the ideas presented here.

I would also like to thank Dr. Wooyoung Kim for her valuable feedback and suggestions that helped me complete this thesis.

I extend sincere thanks to Hajira Rana and Eugene Cheung for their contributions to manuscript review and dataset validation, and to Dr. Brent M. Znosko of Saint Louis University for providing guidance on RNA motif data.

Finally, I am grateful to my family for their enduring support throughout my graduate studies.

DEDICATION

To everyone who taught me that curiosity, not certainty, is the engine of discovery. My high school teachers, college professor and university professors. And most importantly my parents Murugadass and Jamuna packia lakshmi, and my lovely sister Muzhumathi and my little niece Meenu.

Chapter 1

INTRODUCTION

RNA molecules occupy a central position in the molecular biology of life. Unlike DNA, which primarily serves as a stable repository of genetic information, RNA is a dynamic, single-stranded polymer capable of folding into complex three-dimensional architectures that underlie a vast array of cellular functions - from catalysing biochemical reactions as ribozymes, to regulating gene expression through micro-RNAs and riboswitches, to directing protein synthesis as messenger and transfer RNA (2; 3). The structural versatility of RNA is inseparable from its function: the same sequence can adopt different folds to interact with different partners, and small perturbations in three-dimensional structure can abolish or alter biological activity.

Central to RNA folding are secondary structural motifs - recurring three-dimensional arrangements of nucleotides that appear across diverse RNA families. These motifs include hairpin loops, internal loops (both symmetric and asymmetric), bulges, and multi-branch junctions. Each motif class represents a stereotyped departure from ideal Watson–Crick double-stranded helices, creating the bends, kinks, and widened grooves through which RNA molecules recognise one another, bind small-molecule ligands, and assemble into functional machines such as the ribosome (3; 4).

Each panel in Figure 1.1 illustrates a distinct way in which unpaired nucleotides interrupt or cap an RNA double helix. An asymmetric internal loop (a) separates two helices with unpaired residues on both strands but in unequal numbers, for example a 3×1 loop in which one strand contributes AAG and the other a single U between flanking G–C stems. A symmetric internal loop (b) has the same two-helix architecture but with equal counts on each side, such as a 2×2 loop formed by tandem non-canonical residues like GA/AG (often

sheared G·A pairs). A bulge (c) has unpaired nucleotides protruding from only one strand while the opposite strand remains continuously paired, as in a two-nucleotide CU bulge where the helix resumes directly across from it. A hairpin loop (d) is a single stem that folds back on itself and is capped by a run of unpaired nucleotides, the classic case being an ultra-stable UUCG tetraloop or a GNRA loop such as GAAA closing a short stem (e.g., GGAC-UUCG-GUCC).

1.1 Motivation

The importance of RNA structure in disease and therapy has accelerated interest in computational methods for predicting three-dimensional RNA conformation from sequence or from experimental density data. Cryo-EM now routinely resolves macromolecular assemblies at near-atomic resolution, producing volumetric density maps into which atomic models must be fitted or built de-novo. For proteins, this challenge is increasingly well-solved by deep learning pipelines such as DeepTracer (5) and ModelAngelo (6). For RNA, however, the landscape is far less mature: the chemical differences between nucleotides and amino acids, the relatively small number of high-resolution RNA structures in public databases, and the absence of large-scale, motif-resolved training data have all impeded progress.

A key observation motivating this thesis is that RNA secondary structural motifs are not merely topological curiosities - they are the building blocks of RNA tertiary architecture and the primary sites of drug binding and biological regulation (4). Hairpins and internal loops are frequent targets for antibiotics and antiviral compounds (7). If a deep learning model can learn to recognise these motifs directly from cryo-EM density, it gains a structured, biologically meaningful inductive bias that is absent from most existing sequence-only predictors and end-to-end cryo-EM reconstruction methods, which typically do not explicitly model RNA secondary structural motifs as intermediate representations. Making the motif the central object of prediction is also what gives the pipeline a concrete real-world use case: it turns a global cryo-EM map into a set of localised, labelled, drug-relevant structural elements together with a traced backbone, rather than a single monolithic reconstruction. This

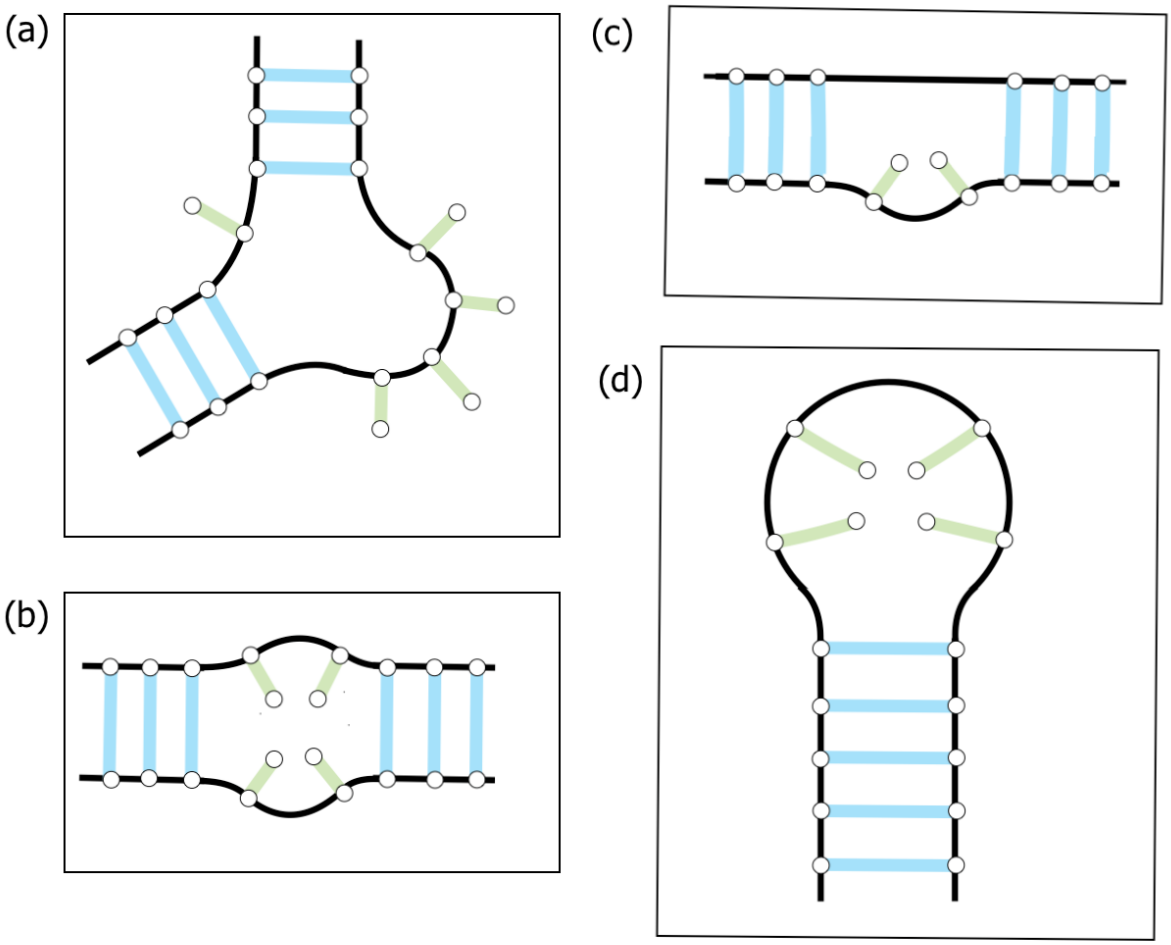


Figure 1.1: Four schematic RNA secondary-structure motifs

The black curve is the sugar phosphate backbone, open circles are nucleotides, blue rungs are canonical Watson–Crick/wobble base pairs forming the stems and green rungs mark unpaired loop nucleotides. (a) shows an asymmetric internal loop, (b) a symmetric internal loop, (c) a bulge and (d) a hairpin loop.

thesis pursues that idea, building a complete pipeline from raw cryo-EM density to backbone atomic coordinates with motif identity as the central intermediate representation.

A further consideration is map resolution. High-resolution cryo-EM maps are the easiest setting for backbone tracing, but they remain comparatively scarce, whereas lower-resolution maps are far more abundant and cheaper to obtain. A practical motif-aware method should therefore aim to remain useful as resolution degrades, and an explicit question addressed in this thesis is how far motif-level interpretation can be pushed toward lower-resolution data and where preprocessing such as density sharpening becomes necessary. The 3D cryo-EM RNA Motif dataset spans 1.5 to 34.0 Å precisely so that this resolution dependence can be measured rather than assumed.

1.2 Problem Statement

The central challenge addressed by this thesis is automated, motif-aware RNA backbone prediction from cryo-EM density maps, which decomposes into two nested sub-problems. The first is *motif localisation*: given a global cryo-EM map that may span hundreds of voxels in each dimension and contain both RNA and protein density, the pipeline must identify the spatial locations at which each type of RNA motif is likely to occur. The second is *backbone prediction*: within each localised, classified region, the pipeline must predict the per-voxel probability that a voxel belongs to the RNA phosphate–sugar backbone, enabling downstream reconstruction of a phosphorus-atom coordinate trace. Solving both sub-problems requires a large, high-quality dataset of motif-resolved cryo-EM density segments as well as models capable of exploiting it, and both are contributions of this thesis.

1.3 Contributions

This thesis makes three primary contributions, matching the three phases described in the abstract.

The first contribution is the 3D cryo-EM RNA Motif dataset (Chapter 3): a large-scale, open-source collection of over one hundred thousand motif-resolved cryo-EM density

segments spanning 25 RNA secondary structural motif classes, each paired with voxel-level backbone, ribose, and nucleobase labels derived from atomic models. The dataset covers a resolution range of 1.5 to 34.0 Å, is validated with cross-correlation fitting scores and Q-score resolvability measures that demonstrate strong map–model agreement across all classes, and is publicly released at https://github.com/DrDongSi/3DEM-RNA-Motif-Dataset and on Zenodo at https://zenodo.org/communities/3dem-rna-motif-dataset (1).

The second contribution is Motif 3D CNN (Chapter 4), a 3D convolutional classifier trained on five consolidated motif classes that reaches a peak validation accuracy of 96.8% and a specificity of 0.948 on high-resolution maps, establishing that cryo-EM density features are discriminative across motif types (1).

The third contribution is an end-to-end inference pipeline (Chapters 5 and 6) that locates motifs in previously unseen global maps by FFT template matching, classifies each candidate with Motif 3D CNN, predicts per-voxel backbone probability with a MONAI UNETR segmentation model, and reconstructs phosphorus-atom PDB files by skeletonisation and chain ordering. The full pipeline was demonstrated end to end on the *Bacillus subtilis* ApdA-stalled ribosomal complex (EMD-18332), where motifs were localised, classified, and traced into a phosphorus-atom backbone visualised in ChimeraX. Across the evaluated RNA samples, backbone probability–density correlations ranged from CC = 0.1219 for a low-resolution hairpin to CC = 0.996 for a high-resolution bulge.

1.4 Thesis Organisation

Chapter 2 provides background on RNA structure, cryo-EM data, and relevant prior work. Chapter 3 describes the construction and validation of the 3D cryo-EM RNA Motif dataset. Chapter 4 presents the motif classification methodology and results. Chapter 5 describes the full inference pipeline for locating motifs in 3D density maps and predicting backbone structure. Chapter 6 reports experimental results on held-out maps. Chapter 7 concludes and outlines future directions.

Chapter 2

BACKGROUND AND RELATED WORK

2.1 RNA Structure and Motifs

RNA is a single-stranded polymer composed of four nucleotide residues - adenosine (A), cytidine (C), guanosine (G), and uridine (U) - linked by phosphodiester bonds through the 2′-hydroxyl-bearing ribose sugar. Unlike DNA, the 2′-OH group restricts the backbone to a characteristic C3′-endo conformation and enables the formation of intra-molecular hydrogen bonds that drive folding. RNA folds first into secondary structure, which is characterised by Watson–Crick and wobble base pairs forming double-stranded A-form helices, and subsequently into tertiary structure through long-range contacts between secondary structural elements as shown in Figure 1.1 (2). Throughout the thesis, motif classes are named by their loop composition. Internal loops are denoted $n{\times}m$, where n and m are the numbers of unpaired nucleotides on each of the two strands of the loop; thus $1{\times}2$ is an asymmetric internal loop with a single unpaired nucleotide opposite two, and $5{\times}5$ is a symmetric internal loop with five unpaired nucleotides on each strand. Bulges are denoted `bulge`n, where n is the number of unpaired nucleotides bulged out from a single strand, so `bulge1` is a single-nucleotide bulge and `bulge5` a five-nucleotide bulge. Hairpins follow the same convention as `hairpin`n, with n the number of unpaired nucleotides in the loop; numbering begins at `hairpin3` because a hairpin loop closed by a single base pair requires a minimum of three nucleotides. This nomenclature is used consistently in this thesis.

2.1.1 Hairpins

A hairpin loop is formed when a single RNA strand folds back on itself, creating a double-stranded helical stem capped by a loop of unpaired nucleotides. Hairpins are among the

most abundant RNA secondary structural motifs and serve as nucleation points for RNA folding, protein recognition sites, and components of ribozymes. In this work, hairpin classes are defined by the number of nucleotides in the closing loop: `hairpin3` through `hairpin7`, and Figure 2.1 shows a sample of hairpin cryo-EM maps overlaid on the atomic model.

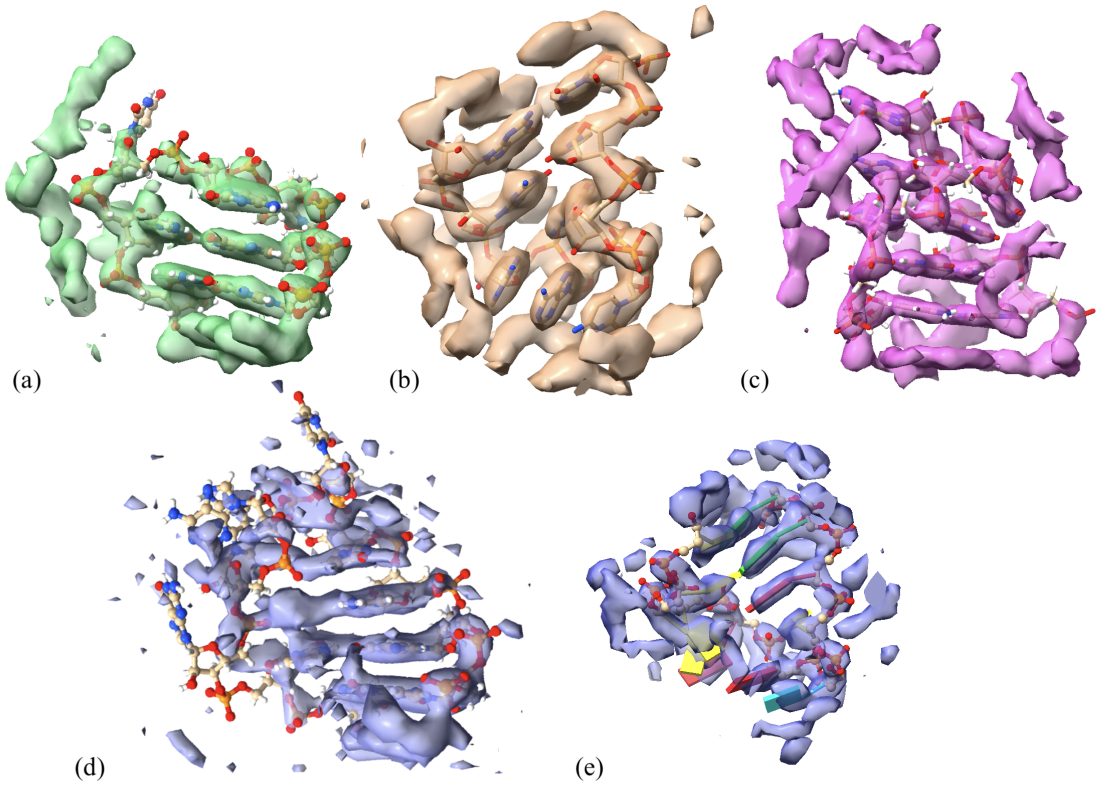


Figure 2.1: **Hairpins motif samples**

The motif type and its respective PDB ID: (a) hairpin3 6S05 – human spliceosome, (b) hairpin4 7PAT – human 80S ribosome, (c) hairpin5 5WF0 – bacterial ribosome, (d) hairpin6 6HCM – bacterial ribosome, and (e) hairpin7 7PD3 – human spliceosome. (1)

2.1.2 Internal Loops

An internal loop arises when two opposing strands of an A-form helix each carry unpaired nucleotides, producing an asymmetric or symmetric bulge within the double-stranded region.

Symmetric internal loops contain equal numbers of unpaired bases on each strand (1×1, 2×2, up to 5×5), while *asymmetric* internal loops carry different counts (1×2, 1×3, up to 4×5). Internal loops alter local helical geometry and mediate many tertiary interactions, including the sarcin-ricin loop of the large ribosomal subunit (3). They are also among the most frequent drug-binding sites in RNA therapeutics (4).

A concrete example appears in the yeast 80S ribosome: residues 140–144 (CCCGA) and 124–128 (UCUGG) form a symmetric internal loop in which only the central G-C base pair is formed; the flanking bases remain unpaired, preventing continuous helical stacking and producing the characteristic kink observed in the density map. Many similar loops are shown in Figure 2.2 and Figure 2.3.

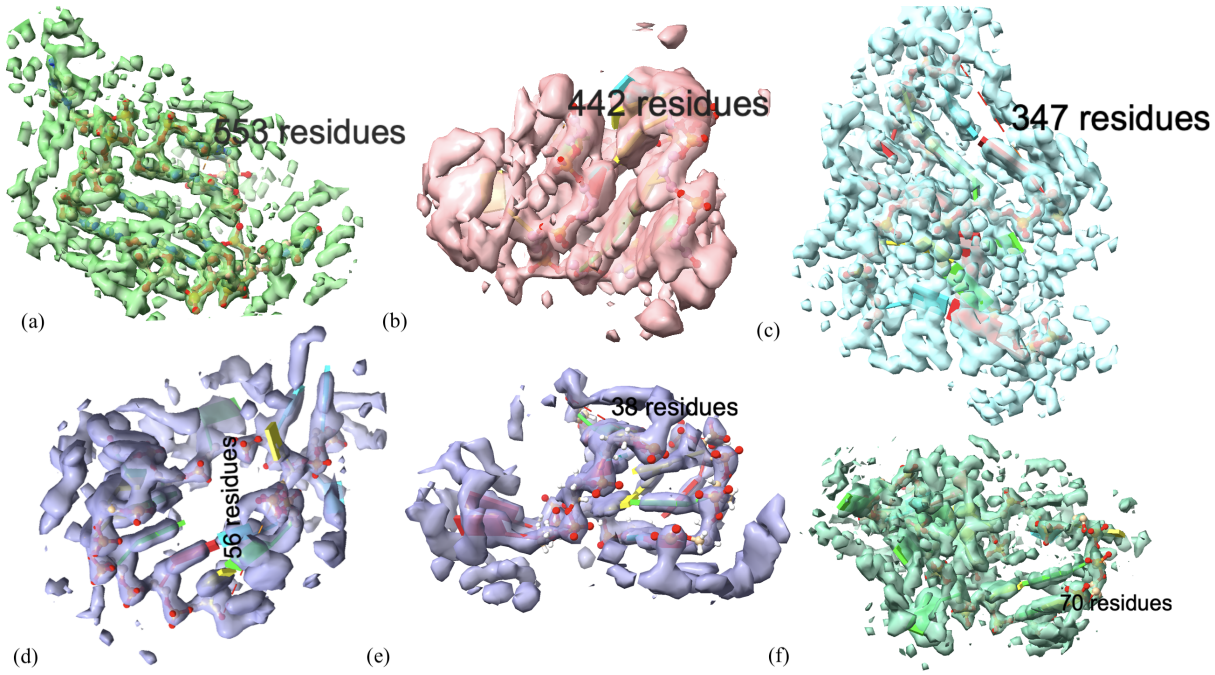


Figure 2.2: **Asymmetric Loop**

The motif type and its respective PDB ID: (a) 1×2 8A3D – bacterial 50S ribosomal subunit, (b) 1×3 7OF5 – large ribosomal subunit (50S), (c) 1×4 6YSI – human mitochondrial large ribosomal subunit (39S mitoribosome), (d) 1×5 6XIQ – SARS-CoV-2 frameshifting RNA pseudoknot, (e) 2×3 6GQ1 – Escherichia coli 50S ribosomal subunit, (f) 2×4 6GC1 – bacterial large ribosomal subunit

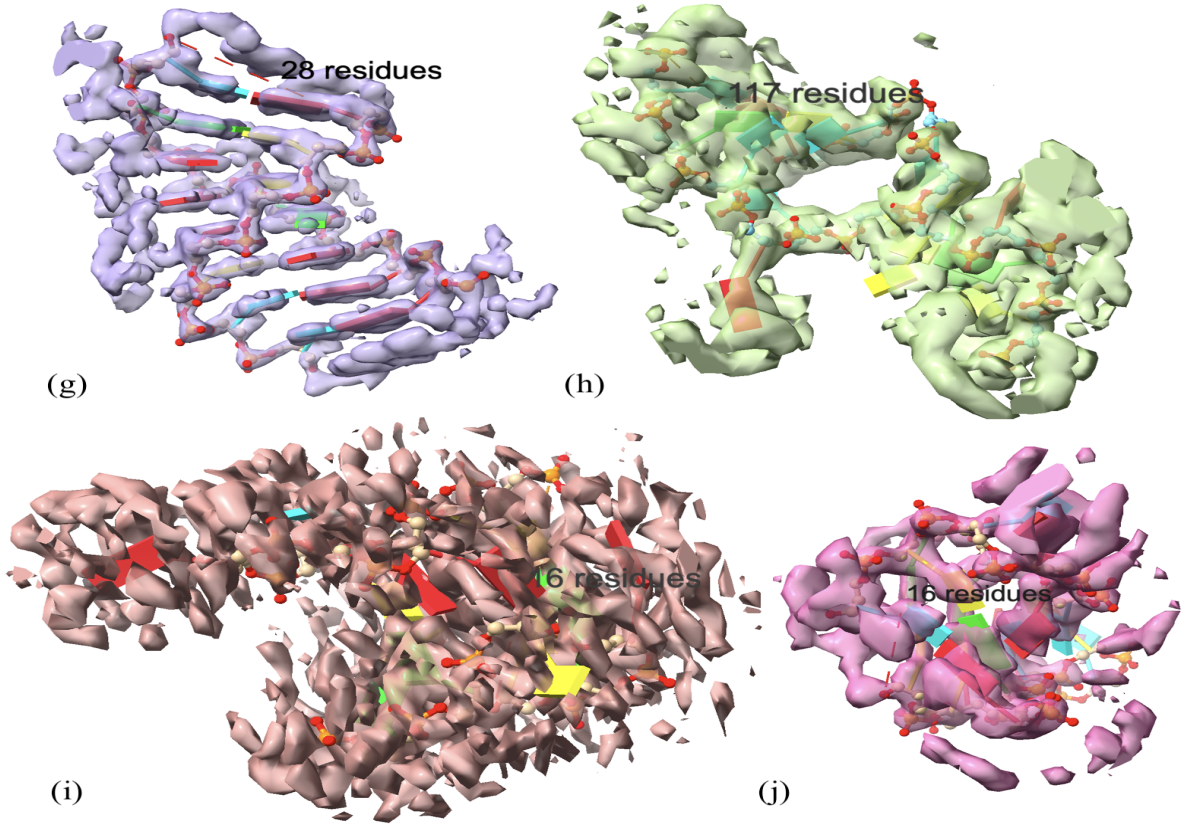


Figure 2.3: **Asymmetric Loop**

The motif type and its respective PDB ID: (g) 2×5 6GQ1 – Escherichia coli 50S ribosomal subunit, (h) 3×4 6GQ1 – Escherichia coli 50S ribosomal subunit, (i) 3×5 7MT7 – human spliceosome, and (j) 4×5 6GQB – Escherichia coli 50S ribosomal subunit.

2.1.3 Bulges

A bulge loop is a special case of an internal loop in which only one strand carries unpaired nucleotides. Bulges range from one to five nucleotides in length in the classes considered here (`bulge1`–`bulge5`) and are notable for making the RNA backbone particularly flexible at the bulge site; some sample bulge motifs are shown in Figure 2.4.

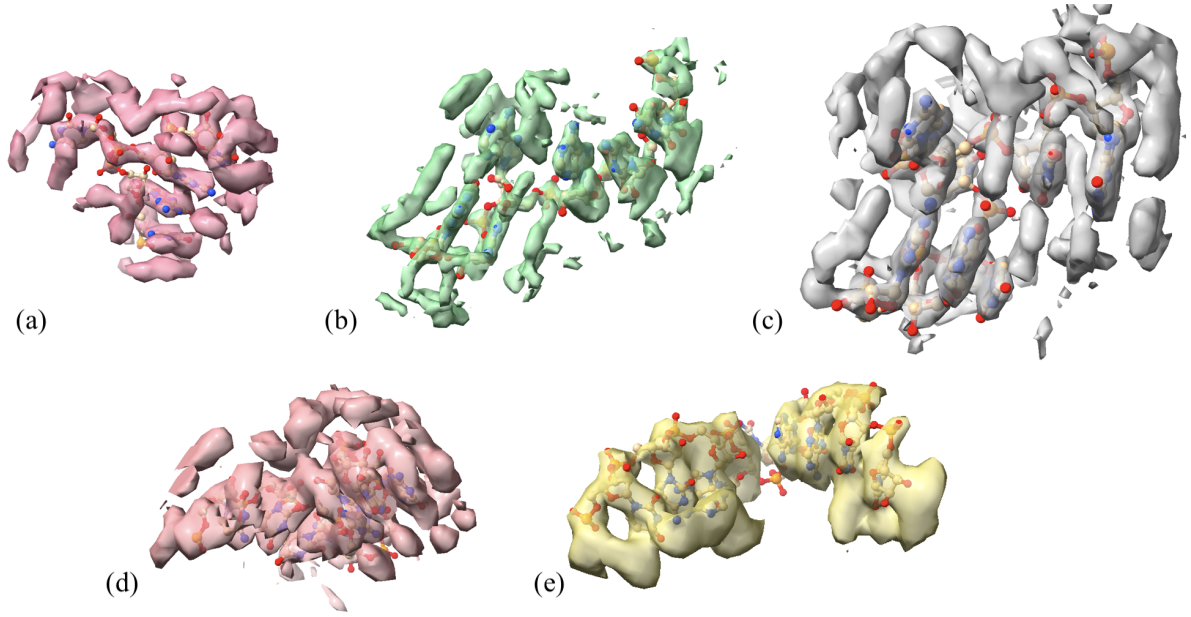


Figure 2.4: **Bulge**

The motif type and its respective PDB ID: (a) bulge1 4V6Y – Tetrahymena group I intron ribozyme, (b) bulge2 4UJE – CRISPR–Cas9–sgRNA–target DNA complex, (c) bulge3 6TC3 – human spliceosome, (d) bulge4 6Y57 – human spliceosome, and (e) bulge5 6ZMO – human spliceosome.

2.2 Cryo-Electron Microscopy Density Maps

Cryogenic electron microscopy (cryo-EM) determines the three-dimensional structure of macromolecules by imaging thousands of single particles embedded in vitreous ice and computationally merging the resulting two-dimensional projections into a volumetric electron density map. Modern cryo-EM routinely achieves resolutions below 3.0 Å for large assemblies, and maps at 2.0–5.0 Å resolution are sufficient to trace nucleic acid backbones and, in many cases, to resolve individual nucleobases (8).

Density maps are stored in MRC format as three-dimensional arrays of floating-point values, with each value representing the averaged electron scattering potential at a given voxel position. The voxel size (typically 0.5–2.0 Å/voxel for near-atomic resolution maps) determines the spatial resolution of the digitised density. Map-model agreement is quantified by

the real-space cross-correlation coefficient (CC), defined as the Pearson correlation between density values sampled from the experimental map and from a model-derived Gaussian map on the same grid (9).

2.3 Related Work

2.3.1 RNA Motif Resources and Sequence-Based Motif Analysis

RNA motif resources describe recurring RNA patterns at several distinct levels of biological organisation – sequence motifs, RNA–protein binding motifs, secondary-structure motifs, and three-dimensional structural motifs. These resources have been central to studying RNA folding, molecular recognition, and functional annotation, but they generally do not provide motif-resolved experimental cryo-EM density.

RNA CoSSMos (10) curates RNA secondary-structure motifs from experimentally resolved atomic models, defining fine-grained classes – hairpin loops, single-nucleotide bulges, and symmetric and asymmetric internal loops – using base-pairing patterns and three-dimensional geometry. It therefore supplies high-quality motif definitions and structural metadata, but is derived from atomic coordinates and contains no volumetric density. The RNA 3D Motif Atlas (11) similarly catalogues recurrent three-dimensional motifs (hairpin loops, internal loops, junctions) identified from PDB structures by geometric similarity, and again provides no localised cryo-EM density.

Other resources are primarily sequence-level. RNAcentral aggregates non-coding RNA sequences and annotations from many expert databases (12), while RBPDB catalogs RNA–protein binding sequence motifs and binding preferences, often as position weight matrices derived from biochemical experiments (13). These are valuable for sequence conservation, RNA family annotation, and RNA–protein recognition, but they neither define three-dimensional structural motifs nor pair atomic coordinates with density maps.

Several motif-identification and classification methods operate on sequences, alignments, secondary structures, or atomic coordinates rather than density. Structure-based search tools

such as FR3D scan resolved RNA structures for recurrent geometric and base-interaction patterns (14); sequence and model-based approaches such as JAR3D and RMDetect infer candidate 3D motifs from sequences, alignments, or known motif models (15; 16); and, more broadly, Rfam and Infernal use covariance models to represent and search RNA families, while CMfinder identifies conserved motifs from sequence data (17; 18; 19). These methods are important for motif discovery and sequence-based classification, but none are designed to classify RNA structural motifs directly from localised experimental cryo-EM density.

Table 2.1 summarises these resources. Only RNA CoSSMos and the RNA 3D Motif Atlas provide explicit structural motifs with atomic examples, and none of the surveyed resources supply motif-level cryo-EM density or voxel-aligned structural labels. This gap motivates a machine-learning-ready resource that links RNA motif identity, atomic structure, and experimental density – the contribution of the 3D cryo-EM RNA Motif dataset.

Table 2.1: Comparison of representative RNA motif and sequence resources with respect to structural annotation and motif-level cryo-EM density availability (1).

Resource / method	**Primary motif or annotation type**	**Structural motifs**	**Atomic 3D coordinates**	**Motif-level cryo-EM density**
RNA CoSSMos	RNA secondary-structure motifs (loops, bulges, internal loops)	Yes	Yes, PDB-derived	No
RNA 3D Motif Atlas	Recurrent RNA 3D motifs (loops, junctions)	Yes	Yes, PDB-derived	No
RNAcentral	RNA sequences and expert database annotations	Limited	No	No
RBPDB	RNA–protein binding sequence motifs and preferences	No	No	No
FR3D / JAR3D / RMDetect	Motif search, annotation, sequence-based 3D recognition	Yes	Yes or model-derived	No
Rfam / Infernal / CMfinder	RNA family / conserved sequence–structure models	Limited	No	No
3D cryo-EM RNA Motif dataset ((1))	RNA structural motifs paired with localised experimental density	Yes	Yes	Yes

2.3.2 Deep Learning for cryo-EM Structure Modelling

DeepTracer (5) introduced an end-to-end pipeline for de-novo protein structure modelling from cryo-EM density maps using a 3D CNN, demonstrating that voxel-level structural labels derived from atomic models could be used to train accurate backbone tracing models. However, DeepTracer was designed exclusively for proteins and does not generalise to RNA due to the different chemical composition and structural organisation of nucleic acids.

CryoREAD (8) is the most closely related prior work to the present thesis in that it targets nucleic acid structure from cryo-EM. CryoREAD uses deep neural networks to predict sugar, phosphate, base, and base-type positions in cryo-EM density maps and then connects identified sugar positions to form the nucleic acid backbone. Tested on maps at 2.0–5.0 Å, it reports backbone recall of 0.856–0.914 and substantially outperforms earlier tools. Crucially, however, CryoREAD does not enumerate or label RNA secondary structural motifs and operates on whole maps rather than motif-centred segments, so it is not a direct baseline for the motif-resolved localisation and classification task studied here; the present pipeline instead treats CryoREAD as a complementary method for the non-motif density regions between detected motifs.

Cryo2Struct (20) and cryo2structData (20) pioneered the construction of large, openly accessible datasets linking cryo-EM maps with voxel- level structural annotations for proteins, demonstrating the value of such resources for training generalizable models. The multi-tiered labelling schema (atom-level, Cα-only, residue-type, secondary structure) directly informed the labelling strategy adopted in the 3D cryo-EM RNA Motif dataset.

Finally, UNETR (21) combines a Vision Transformer (ViT) encoder with a CNN U-Net decoder for 3D medical image segmentation, arguing that the global self-attention mechanism of the transformer is better suited to long-range spatial dependencies than convolutional filters alone. The ViT encoder's global attention is particularly well-suited to backbone tracing, where the phosphate chain is a long-range structure spanning the full motif cube. The MONAI production implementation of UNETR is used directly in this work (22).

CryoDataBot (23) offers a scalable automated pipeline for generating labelled cryo-EM datasets from EMDB and PDB, emphasising reproducibility and integration with WorkflowHub. It provides a useful complementary perspective on large-scale cryo-EM data curation. However, the framework is designed primarily for general biomolecular dataset generation rather than RNA-specific structural analysis. It does not explicitly identify or segment RNA secondary structural motifs such as hairpins, internal loops, bulges, junctions, or pseudoknots, which are fundamental building blocks of RNA structure and function. Furthermore, CryoDataBot does not integrate RNA-specific annotation resources such as RNA CoSSMos or perform motif-level extraction by combining atomic coordinates with cryo-EM density maps. Consequently, although it is well suited for automated data acquisition and preprocessing, it cannot directly produce the motif-centric datasets required for training deep learning models for RNA motif classification or atomic structure prediction.

Chapter 3

THE 3DEM-RNA-MOTIF DATASET

3.1 Overview

The 3D cryo-EM RNA Motif dataset is the primary enabling resource for all downstream deep learning in this thesis. It contains over one hundred thousand motif-resolved cryo-EM density and atomic structure segments spanning 25 classes of RNA secondary structural motifs, each paired with voxel-level structural labels derived from atomic models. The dataset covers a resolution range of 1.5 to 34.0 Å, encompassing both near-atomic and lower-resolution density maps relevant to RNA modelling tasks. Source code, the fully automated construction pipeline, and the benchmark dataset are publicly available at https://github.com/DrDongSi/3DEM-RNA-Motif-Dataset and on Zenodo at https://zenodo.org/communities/3dem-rna-motif-dataset (1). The construction pipeline is fully scripted and driven from the public PDB-to-EMDB cross-reference and the RNA CoSSMos annotations, so that the release can be regenerated end-to-end without manual curation.

3.2 Data Sources

The dataset integrates three complementary types of experimental data. cryo-EM density maps are obtained from the Electron Microscopy Data Bank (EMDB) (24), which archives three-dimensional macromolecular volume maps produced by cryo-EM across a wide resolution range and supplies the volumetric density required to segment and label RNA structural regions. Atomic-level RNA structures are obtained from the RCSB Protein Data Bank (25), whose coordinate files provide the three-dimensional atomic positions used to define motif boundaries and generate voxel-level labels. Motif annotations are sourced from RNA CoSSMos (Context of Secondary Structure Motifs) (10), a specialised database curated by

the Znosko Lab that catalogs secondary structural motifs from RNA entries in the PDB together with detailed metadata such as hydrogen-bonding patterns, glycosidic-bond conformations, and nucleotide identities; these annotations provide the ground-truth class labels for all 25 motif types.

3.3 Motif Classes

The dataset targets four biologically important motif categories, expanded into 25 fine-grained classes based on loop size and symmetry.

Hairpins and bulges are of particular interest as frequent drug-binding sites, making them attractive targets for structure-based drug discovery. Symmetric loops tend to exhibit more regular geometries, which are advantageous for computational modelling, whereas asymmetric loops are the most common class by raw instance count. The complete per-class instance counts are reported in Appendix A.1; the resource is naturally imbalanced, with hairpins and bulges dominating the totals, and this imbalance is addressed explicitly in the class-balanced blind-test protocol of Chapter 6.

3.3.1 Retrieval

For each motif instance annotated in RNA CoSSMos, the corresponding PDB identifier is used to retrieve the atomic model from the RCSB PDB and the associated density map from the EMDB. Matching between PDB and EMDB entries is performed using the official PDB-to-EMDB cross-reference table.

3.3.2 Segmentation

Segmentation isolates motif-specific density regions from the full global cryo-EM map based on atomic structural annotations. For each RNA structure, as shown in Figure 3.1, there can be multiple motifs present in the global density map; motif residues are first identified in the atomic model using the RNA CoSSMos annotations, and their coordinates define a

spatial neighbourhood around the motif. The cryo-EM density within a fixed 5.0 Å radial distance of any motif atom is extracted to ensure that local structural context is preserved, including the flanking Watson–Crick helical stems. The 5.0 Å radius is chosen to be large enough to retain both flanking stems – which anchor the motif geometry – while excluding neighbouring, unrelated density. Where a residue could in principle be assigned to more than one overlapping annotation, the RNA CoSSMos class label is taken as authoritative and the segment is extracted for that annotated motif, so each extracted segment carries a single, unambiguous class.

After extraction, the density segment is cropped to its minimal bounding box by identifying all non-zero density voxels and computing the smallest three-dimensional bounding box that encloses them. This removes empty background regions while preserving local structural context, and produces compact, motif-specific cryo-EM density segments.

The density maps are then resampled to a constant isotropic voxel size of 0.5 Å/voxel using trilinear interpolation. This standardisation is essential because raw EMDB maps vary in voxel size from approximately 0.5 to 2.5 Å/voxel depending on the reconstruction protocol and the microscope's calibrated pixel size. Uniform resampling ensures that all input features are in the same spatial reference frame for model training.

3.3.3 Labelling

Voxel-level structural labels are generated by projecting atomic-model coordinates into the resampled density grid, following the procedure formalised in Algorithm 1 and illustrated in Figure 3.2 (1). After loading the resampled cryo-EM map and its corresponding PDB model, an empty label array is initialised with the same dimensions as the density grid. The atomic model is then traversed chain by chain and residue by residue, and each atom is assigned to one of three annotation layers: a *backbone* layer comprising the phosphate (P) and the O3′/O5′ bridging oxygens of the phosphodiester backbone; a *ribose* layer comprising the C1′–C5′ ring atoms of the sugar; and a *nucleobase* layer comprising the heavy atoms of the purine or pyrimidine base. Each atom's Cartesian coordinates are converted to voxel indices using

the affine transform stored in the MRC header, and every voxel within a 3.0 Å Euclidean labelling radius of the atom receives the corresponding label. The 3.0 Å radius reflects the blurred, averaged nature of experimental cryo-EM density: it is large enough to cover the density attributable to a single atom at near-atomic resolution while remaining small enough to avoid systematically merging adjacent atoms of different classes. Where more than one atom class falls within the labelling radius of a voxel, the label of the nearest atom is used, ensuring a biologically consistent assignment; the residual label noise from this projection is therefore concentrated at class boundaries and grows as resolution coarsens, which the quality-validation metrics below make explicit. The completed label array is written as a secondary MRC file with the same dimensions as the density map, using the numeric encoding 0 = background, 1 = backbone, 2 = ribose, 3 = nucleobase. This compact representation keeps label file sizes manageable while encoding the full multi-class annotation.

Medical Research Council (MRC) density map files are labelled by projecting atomic structural information into the three-dimensional voxel grid of the cryo-EM map, as shown in Figure 3.1. The result of labelling is a set of individual labelled density map files, one of which is shown in Figure 3.3.

3.3.4 Density Normalisation

Prior to labelling and model training, each density map is normalised using percentile-based thresholding: the top 5% of non-zero density voxels are retained (95th percentile threshold), suppressing background noise and outlier high-density voxels. The remaining density values are then z-score normalised (zero mean, unit variance) within each segment to ensure consistent intensity scales across maps acquired at different resolutions or with different microscopes (1).

3.4 Quality Validation

The biological validity of the segmented dataset is assessed using two complementary map-model agreement metrics. Because the dataset draws heavily on ribosomal structures, these

intrinsic map–model metrics – rather than an external predictor – are the primary evidence that the segments are biologically correct, and they are reported both per motif class and stratified by resolution so that the reader can judge validity at the map quality relevant to their own application.

3.4.1 Cross-Correlation Fitting Score

Cross-correlation was computed using the PHENIX tool `phenix.map_model_cc` (26), which applies a mean-subtracted, normalized formulation suitable for cryo-EM maps. For an experimental map ρ_1 and a model-derived map ρ_2, the real-space correlation coefficient is defined as

$$N = \sum_{\mathbf{n}} \big(\rho_1(\mathbf{n}) - \langle\rho_1\rangle\big)\big(\rho_2(\mathbf{n}) - \langle\rho_2\rangle\big) \tag{3.1}$$

$$D = \left[\sum_{\mathbf{n}} \big(\rho_1(\mathbf{n}) - \langle\rho_1\rangle\big)^2\right]^{1/2} \left[\sum_{\mathbf{n}} \big(\rho_2(\mathbf{n}) - \langle\rho_2\rangle\big)^2\right]^{1/2} \tag{3.2}$$

$$\mathrm{CC}(\rho_1, \rho_2) = \frac{N}{D} \tag{3.3}$$

where $\rho_1(\mathbf{n})$ and $\rho_2(\mathbf{n})$ are the experimental and model-derived density values at grid point $\mathbf{n}$, respectively. To obtain motif-level assessments, correlations were calculated within a molecular mask surrounding each extracted motif. The resulting $\mathrm{CC}_{\mathrm{mask}}$ reduces contributions from solvent, background, and unrelated neighboring density, and is widely used for local cryo-EM map–model validation (27; 28; 1).

$\mathrm{CC}_{\mathrm{mask}}$ summary statistics were computed from valid fit values. Blank entries and samples with all four correlation measures equal to zero were excluded from the statistics but retained in the reported segmentation counts (1).

The cross-correlation results for all motif classes are reported in Section 6.1. In summary, hairpin motifs yield average CC values of 0.79 ± 0.06, indicating reliable density–model agree-

ment across the five size classes, while bulge motifs yield slightly more variable agreement at 0.76 ± 0.08 owing to the irregular backbone geometry at bulge sites. Symmetric and asymmetric internal loops yield 0.81 ± 0.05 and 0.74 ± 0.09 respectively, with symmetric loops showing the best agreement, consistent with their regular helical geometry.

3.4.2 Q-Score Resolvability

The Q-score (29) measures the resolvability of individual atoms by comparing the map value profile around each atom to the Gaussian profile expected for a well-resolved atom at that resolution. Atomic resolvability was further evaluated using Q-scores, which quantify the agreement between individual atoms in the atomic model and the surrounding experimental cryo-EM density (29). Q-scores compare the atom-centered density profile with an idealized Gaussian reference profile and range from -1 to 1, with higher values indicating stronger atom-level support from the density. Q-scores were computed using the ChimeraX implementation described by (29). Because Q-score values were generated through a separately processed validation workflow, they are reported as complementary atom-level validation results rather than as paired measurements with the CC_{mask} subset.

3.5 Dataset Statistics and Splits

The high resolution instances 1.5 - 2.8 Å from the dataset is partitioned into training (80%), validation (10%), and held-out test (10%) splits using sequence-cluster-pre-filtered random sampling to prevent data leakage between splits. The train/validation/test split was rebuilt from a 1094/62/121-structure partition to ensure genuine independent evaluation on the test set. Clustering by sequence before sampling is what prevents near-duplicate ribosomal structures from appearing on both sides of a split, and therefore limits the extent to which the abundance of closely related ribosomal entries can inflate the reported accuracy.

Label files are stored in the directory structure, in the 3D cryo-EM RNA Motif dataset as `Labels/output/SET{N}/{subtype}/backbone_label_{stem}.mrc`, where the stem matches the corresponding density MRC filename. Training and validation manifests open sourced

through the Github project repository through the files (`training_pairs.csv`, `val_pairs.csv`), which are generated by the splitting framework as described in Table 3.1.

Table 3.1: Training, validation, and blind-test sample counts for the baseline 3D CNN motif classifier. Five models were trained using fold-specific training/validation splits and then evaluated on the blind test set.

Fold	**Training samples**	**Validation samples**	**Blind test samples**
1	1,799	450	438
2	1,799	455	438
3	1,799	460	438
4	1,803	460	438
5	1,803	460	438

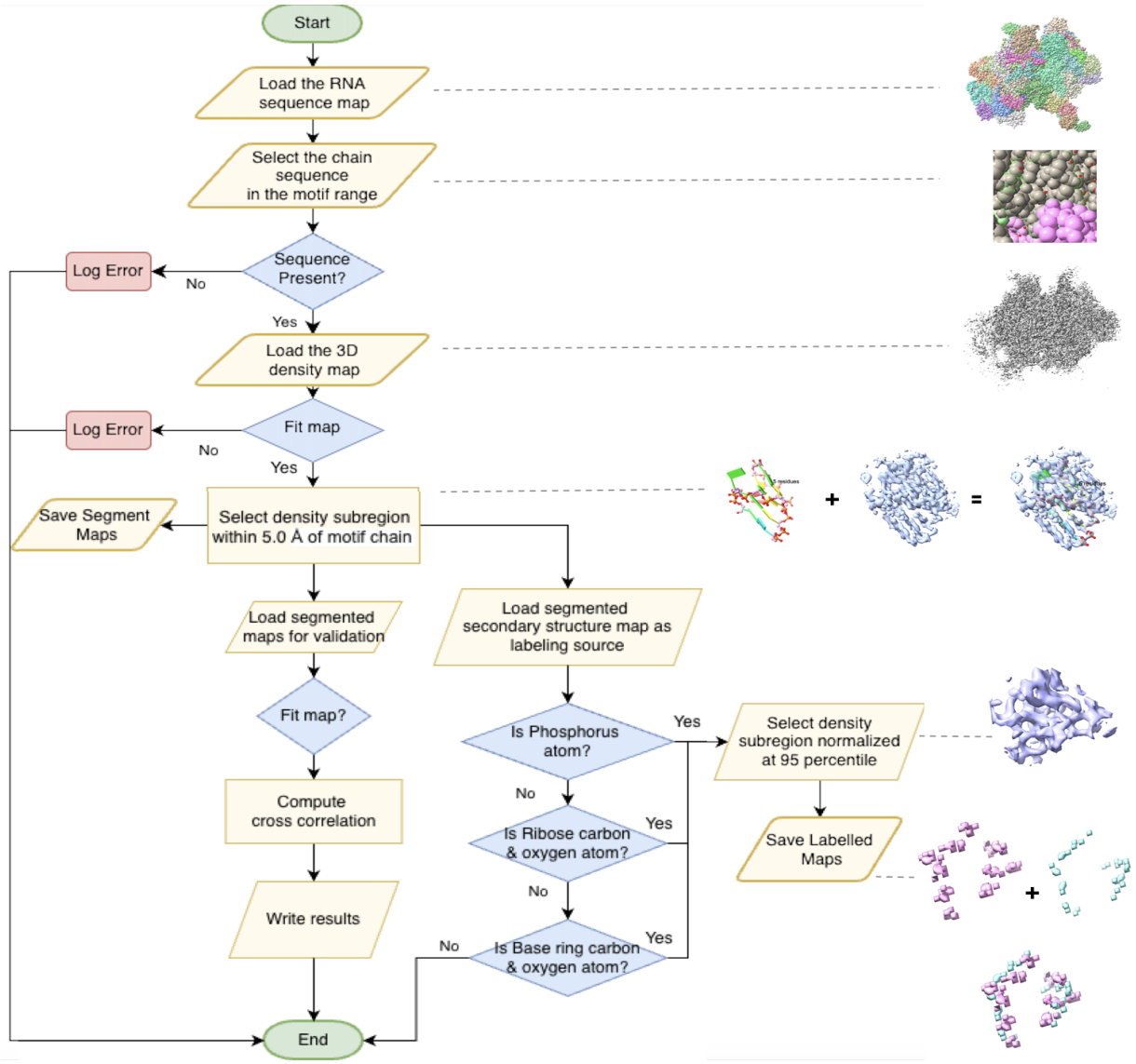


Figure 3.1: **Motif segmentation and labeling**

The flowchart depicting the identification and segmentation of motifs from within the larger cryo-EM density maps. (1)

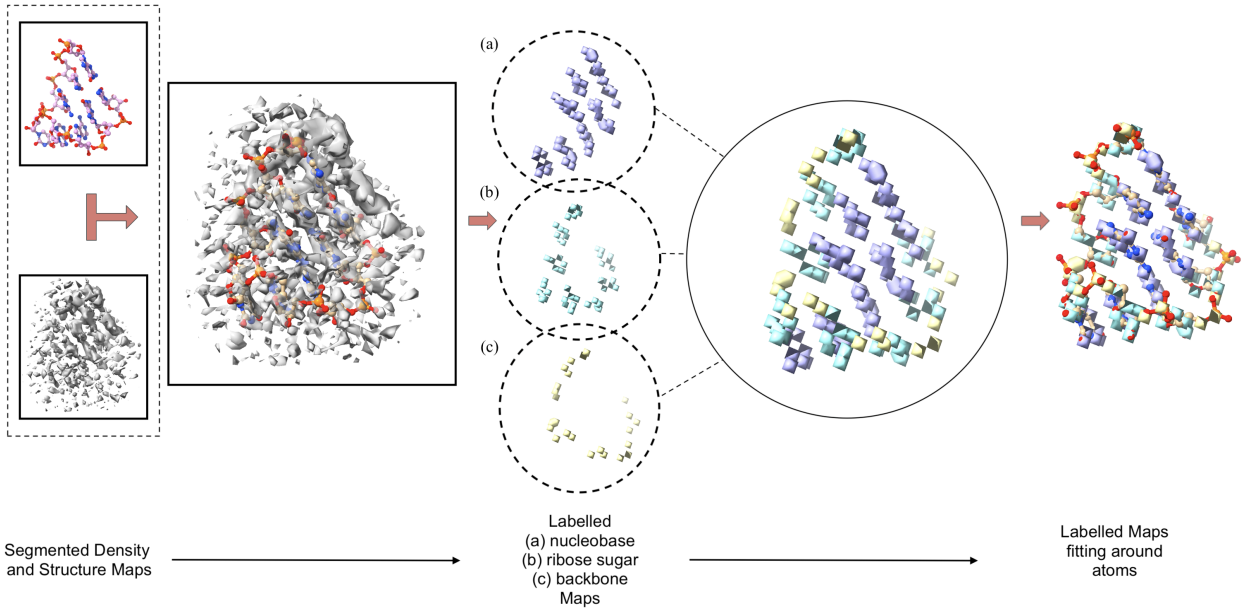


Figure 3.2: **Labelling a segmented motif**

Segmented hairpin with four residues at the location 1896 - 1903 chain positions in Chain A of B *subtilis ApdA-stalled* motif is used to generate labels of ribose sugar, nucleobase and the backbone. (1)

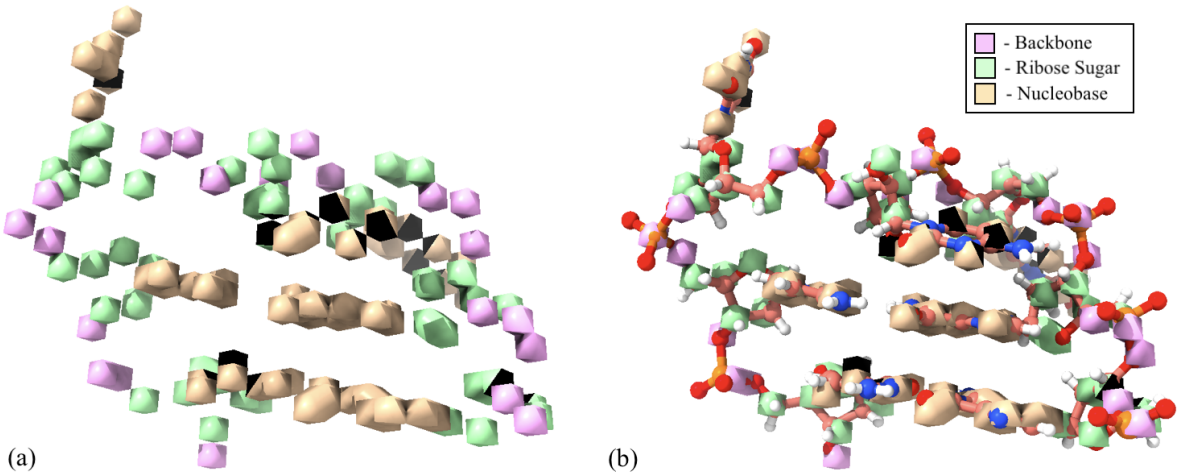


Figure 3.3: **Labeled RNA segment**

Segment of RNA with PDB ID 6PJ6 segmented at the residue chain sequence 85-G 86-G 87-U 88-C 89-U 90-C 91-C to form a hairpin with three nucleotides. (a) The backbone, ribose and nucleobase voxels are labeled for the segmented motif region. (b) The labeled backbone, when fitted onto the corresponding segmented atomic model, demonstrates accurate spatial overlap between the predicted labels and the underlying atomic structure.

Chapter 4

MOTIF CLASSIFICATION

4.1 Overview

Motif classification serves two purposes in this thesis. First, it validates the 3D cryo-EM RNA Motif dataset: for the labelled density segments to be internally consistent and biologically meaningful, a model should be able to learn discriminative features that separate the motif classes. Second, the trained classifier (Motif 3D CNN) is deployed as a core component of the inference pipeline described in Chapter 5, where it filters candidate locations generated by template matching and assigns motif class labels to guide backbone prediction. Fig. 4.1 shows a sample density map with the four classes of motifs which can later be used for localised backbone prediction.

4.2 Consolidated Class Schema

The 25 fine-grained motif classes of the 3D cryo-EM RNA Motif dataset are consolidated into five trainable classes for the classifier, as summarised in Table 4.1.

When the 25-class formulation was initially tested, the model struggled to converge due to the high geometric similarity between adjacent size classes (e.g., `2x2` vs. `3x3` symmetric loops) and the severe scarcity of some fine-grained classes (for example `3x5` and `bulge5`, with only tens of instances each). Consolidating to five classes substantially improved classification stability and enabled the near-state-of-the-art accuracy reported below; the fine-grained schema is retained in the dataset for future work with larger, class-balanced training sets and higher-capacity architectures.

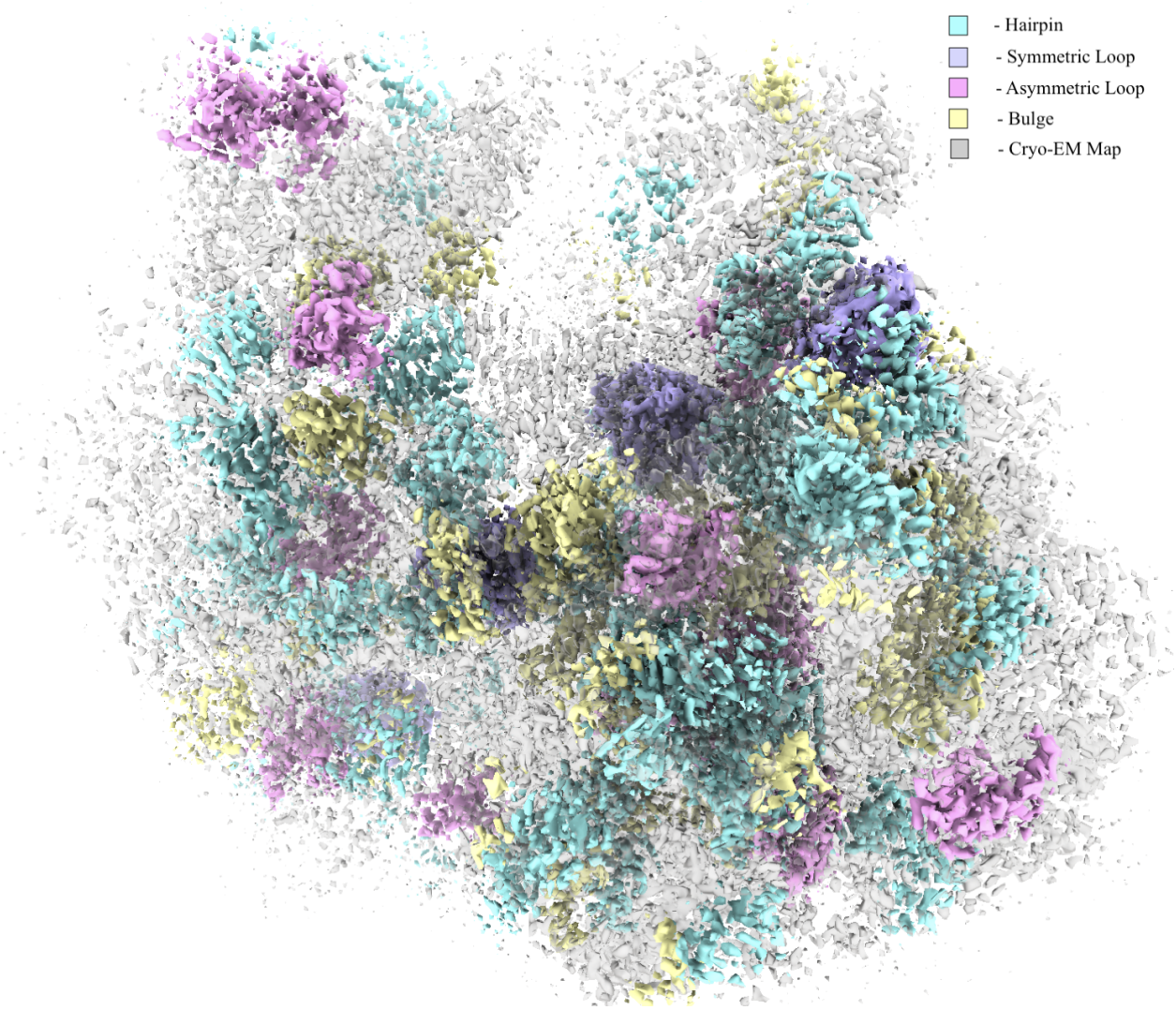


Figure 4.1: **Motif-resolved RNA cryo-EM density maps**

Representative motif-level cryo-EM density segments from the *Bacillus subtilis* ribosomal complex (EMDB ID: EMD-18332), colored by RNA secondary structural class. Hairpin loops (cyan), symmetric internal loops (lavender), asymmetric internal loops (pink), and bulges (yellow) are shown. Each segment corresponds to a localized RNA structural motif extracted from the full experimental density map and aligned with atomic-level annotations.

Table 4.1: Consolidated motif classes used for training Motif 3D CNN. The Subtypes column lists all 25 fine-grained classes that make up each consolidated class.

Class	**Subtypes**	**Notes**
Symmetric loop	1x1, 2x2, 3x3, 4x4, 5x5	Most structurally conserved; regular geometry
Asymmetric loop	1x2, 1x3, 1x4, 1x5, 2x3, 2x4, 2x5, 3x4, 3x5, 4x5	Highest instance count; most geometrically diverse
Hairpin	hairpin3, hairpin4, hairpin5, hairpin6, hairpin7	Largest subgroup by count
Bulge	bulge1, bulge2, bulge3, bulge4, bulge5	Most size-variable class
Unknown	–	Background or unclassifiable regions

4.3 Motif 3D CNN Architecture

Motif 3D CNN is a 3D convolutional neural network that takes as input a pair of volumes: (1) a normalised density map cube and (2) its corresponding backbone label map. The model extracts voxel-level features from the labelled positions within the density and passes them through successive 3D convolutional layers with batch normalisation and ReLU activations, followed by a fully connected classification head with a softmax output over the five classes (1).

A 3D CNN was chosen for the classification stage in preference to a ResNet-style deep residual network or a full 3D transformer for two reasons. First, it is parameter-efficient and trains stably on the moderate number of high-resolution motif segments available, whereas higher-capacity models tend to overfit or fail to converge at this data scale. Second, the features that distinguish motif classes are largely short-range and locally textured – the presence, size, and symmetry of a loop void – which convolutional filters capture directly; the transformer capacity is instead reserved for the backbone-segmentation stage of Chapter 5, where long-range connectivity along the phosphate chain is essential. Because the classifier consumes the paired backbone label channel and is trained on resolution-stratified data, its

decisions are anchored to structural geometry rather than to a single resolution regime, an effect examined directly in the resolution-stratified analysis of Section 6.2.2.

The main architectural choices are as follows. The classifier operates on $32 \times 32 \times 32$ voxel input cubes. At the 1.0 Å/voxel inference sampling this corresponds to a 32 Å field of view, which captures a motif's loop together with its immediately flanking helical stems; the full spatial extent of the largest motifs, which can reach roughly 72 Å, is instead accommodated by the 96^3 (96 Å) field of view used by the backbone-segmentation stage in Chapter 5. To keep the physical field of view fixed across training and inference, training patches are extracted as 64^3 cubes at 0.5 Å/voxel while inference cubes are resampled to 32^3 at 1.0 Å/voxel, so that both span the same 32 Å physical extent (Figure 4.2). The internal compression dimension is set to 32 following ablation experiments, since a value of 8 led to mode collapse during training. Training uses a cross-entropy loss for multi-class classification with the Adam optimiser, and the filtered high-resolution dataset is split 80/20 for training and validation, with the best checkpoint selected by validation accuracy.

4.3.1 Classifier Evaluation Metrics

Because the motif classes are imbalanced, overall accuracy alone can be misleading, and the classifier is therefore assessed with sensitivity, specificity, and their geometric mean. Sensitivity (recall, or true-positive rate) measures the proportion of actual positive samples that are correctly identified and is high when the model misses few positives:

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \tag{4.1}$$

Specificity (true-negative rate) measures the proportion of actual negatives correctly rejected and is high when the model produces few false positives:

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{4.2}$$

Here TP, TN, FP, and FN denote true positives, true negatives, false positives, and false

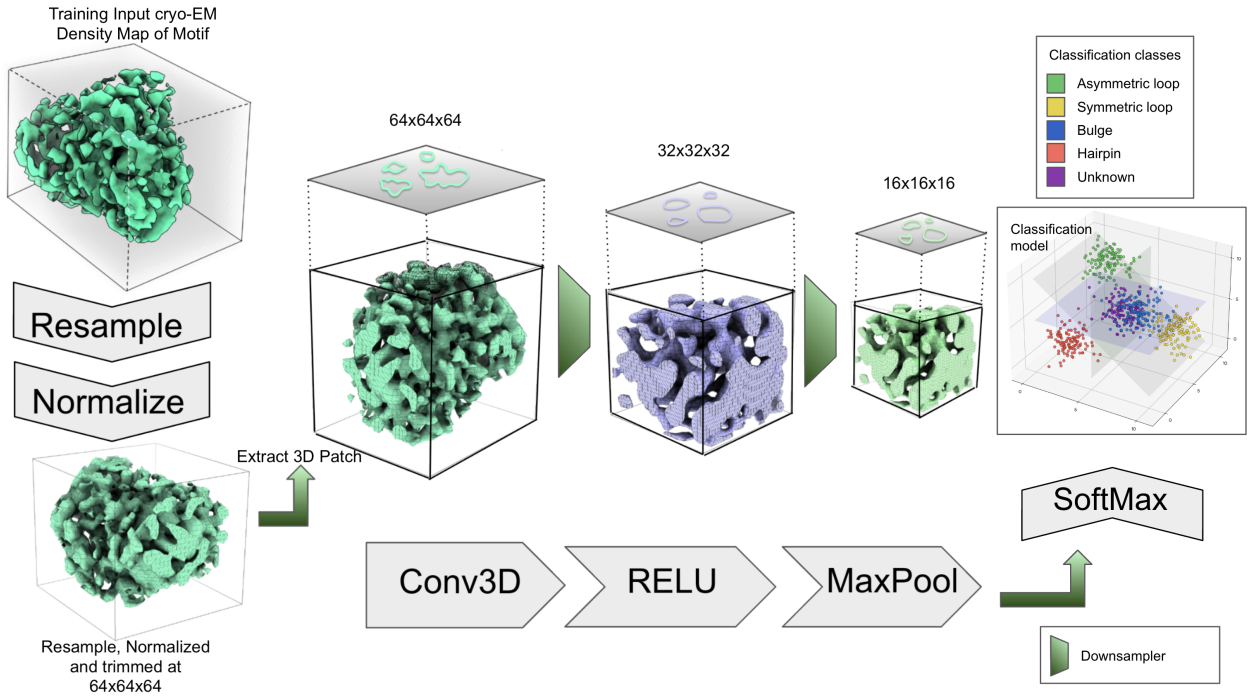


Figure 4.2: **3D CNN Classifier Workflow**

Takes segmented density maps as input and builds a classification model that classifies the known motif data into motif classes. (1)

negatives respectively. The geometric mean (G-mean) combines sensitivity and specificity into a single balanced score:

$$\text{G-mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}}. \tag{4.3}$$

Because it is a geometric rather than an arithmetic mean, a high G-mean requires the classifier to perform well on both the positive and the negative class simultaneously: a low value on either metric drives the G-mean down. Sensitivity and specificity thus provide complementary information – a model with high sensitivity but low specificity detects most positives at the cost of many false positives, and vice versa – and their geometric mean is a compact summary that is particularly informative for the imbalanced motif classes, where a classifier that simply predicts the majority class can achieve high accuracy while failing on

minority classes. These metrics are reported per class and as a macro-average in Table 6.3.

4.3.2 Confusion Matrix

A confusion matrix is a tabular representation used to evaluate a classification model by comparing the predicted class labels with the ground-truth labels; it summarises the number of correct and incorrect predictions and is the basis for computing accuracy, precision, recall (sensitivity), and specificity. For a binary problem it has four entries – true positives, false negatives, false positives, and true negatives – arranged as in Table 4.2.

Table 4.2: Confusion matrix for binary classification.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

The confusion matrix provides a detailed breakdown of classification performance beyond overall accuracy, making it possible to see whether a model tends to miss positives (high false negatives) or to over-predict them (high false positives). In the multi-class setting of RNA motif classification the matrix is extended to an $N \times N$ table in which each row is a true class and each column a predicted class; the diagonal entries are correct classifications and the off-diagonal entries reveal which motif classes are confused, typically owing to structural similarity or class imbalance. The same TP, TN, FP, and FN counts underlie the sensitivity, specificity, and G-mean defined above, as well as precision and accuracy, so the confusion matrix serves as a single, comprehensive summary of classifier performance that is widely used in medical imaging, bioinformatics, and structural biology.

Chapter 5

MOTIF LOCALISATION AND BACKBONE PREDICTION IN 3D DENSITY MAPS

5.1 System Architecture

The end-to-end inference pipeline translates a global cryo-EM density file into a set of annotated motif locations and then to a phosphorus atom backbone trace PDB files. The pipeline comprises five sequential stages, summarised in Table 5.1.

Table 5.1: End-to-end pipeline stages for motif localisation and backbone prediction.

Stage	**Description**
Template Build	We build a labelled density cubes per class to produce a 32^3 mean templates at 0.5 Å/voxel.
Template Matching	FFT cross-correlation of each class template against the global map; non-maximum suppression (NMS) with radius 12 Å and CC threshold 0.25 yielding candidate motif locations.
RNA Density Check	Three density-based checks filter obvious false positives without a GPU forward pass: (1) mean cube density $> 50\%$ of global background mean; (2) internal density variation std/mean > 0.3; (3) $< 40\%$ of voxels above half-maximum.
Classification	Motif 3D CNN (fold5Model.pth) classifies each surviving 32^3 cube into one of five motif classes; detections with confidence < 0.50 or labelled "unknown" are discarded.
Backbone Prediction	MONAI UNETR processes each 96^3 cube and outputs a per-voxel backbone probability map.
Post-processing	Thresholding to P-atom placement via Gaussian filling, closing, skeletonisation, and chain ordering.

Software configuration is summarised in Table 5.2.

Table 5.2: Software and hardware configuration for the inference pipeline.

Category	**Details**
Primary language	Python 3.12
Deep learning	PyTorch 2.x, MONAI 1.3 (UNETR, DiceCELoss, sliding-window)
Scientific libraries	NumPy, SciPy (FFT, ndimage, spatial), mrcfile
Visualisation	UCSF ChimeraX 1.9 (map + PDB overlay)
Compute	NVIDIA GPU (CUDA), CUDA_VISIBLE_DEVICES=2

5.2 Template Construction

For each of the four motif classes, a representative mean density template is constructed from the 3D cryo-EM RNA Motif training set, as described in Algorithm 2. All density MRC cubes for a class are read, resampled to a common voxel size of 1.0 Å/voxel, z-score normalised, and centre-cropped to 32^3 voxels; an element-wise sum is then accumulated across all cubes and divided by the count to give the mean density template.

No rotational alignment is performed between instances; the resulting template encodes the characteristic radial density profile of the motif class, which is sufficient for cross-correlation discrimination even without orientation alignment. Despite this simplification, templates for different motif classes have sufficiently distinct radial density profiles that cross-correlation scores can separate them above the chosen threshold. This omission is a deliberate trade-off between localisation precision and computational cost, and the precision it sacrifices – especially for the orientation-sensitive asymmetric loops – is quantified in Section 6.3 and revisited as future work in Chapter 7.

5.3 FFT Template Matching

Template matching is used as the first localisation stage for three reasons. First, a global cryo-EM map may contain hundreds of millions of voxels, and an exhaustive per-voxel forward pass of the CNN or UNETR would be prohibitively expensive; FFT cross-correlation reduces the naïve $\mathcal{O}(N^2)$ sliding-window search to $\mathcal{O}(N \log N)$ and requires no GPU. Second, it acts

as a class-aware filter that proposes a small set of candidate locations for the expensive downstream models rather than requiring them to scan the entire volume. Third, the radial density profiles of the RNA motif classes are sufficiently distinct – a central loop void for hairpins, an asymmetric deficit for internal loops, and a single-sided kink for bulges – that even orientation-agnostic mean templates separate motif density from the protein background above the chosen correlation threshold. Template matching therefore trades some precision, recovered by the subsequent density checks and CNN classification, for a very large reduction in the search space.

Given a global density map f (which may contain hundreds of millions of voxels for a large ribosome dataset) and a 32^3 class template g, the 3D cross-correlation is computed as:

$$\mathrm{CC}(\mathbf{x}) = \mathrm{IFFT}\big(\mathrm{FFT}(f) \cdot \overline{\mathrm{FFT}(g)}\big) \tag{5.1}$$

followed by normalisation:

$$\widetilde{\mathrm{CC}}(\mathbf{x}) = \frac{\mathrm{CC}(\mathbf{x})}{\sigma_f \cdot \sigma_g} \tag{5.2}$$

where σ_f and σ_g are the standard deviations of the map and template densities respectively. The result is a 3D score map of the same dimensions as the input map, where the value at position $\mathbf{x}$ reflects the degree of similarity between the template and the local density centred at $\mathbf{x}$.

Candidate peaks are identified by thresholding $\widetilde{\mathrm{CC}}$ at 0.25. Non-maximum suppression (NMS) with a radius of 12 Å (approximately one motif width) is then applied to remove duplicate detections arising from the same motif instance. The CC threshold and NMS radius were selected on held-out maps as the operating point that recovered the RNA-rich regions while keeping the candidate count tractable for the downstream stages: lowering the CC threshold or the NMS radius increases recall at the cost of many more protein-background candidates, while raising them suppresses genuine, closely spaced motifs, so the chosen values favour recall and delegate precision to the density checks and classifier. On a test sample of the EMD-18332 ribosome map, this stage yields approximately 186 candidate

locations per class template run.

5.4 Template Matching Evaluation

Because template matching determines which regions are ever passed to the classifier and backbone model, the quality of the complete pipeline is bounded by the quality of this stage, and it is therefore evaluated as a localisation task in its own right. Detections are compared against the RNA CoSSMos motif centroids annotated for the same map. A detection is counted as a true positive if its centre lies within a matching radius r_{match} of an annotated motif centroid and is assigned, at most once, to the nearest such centroid; unmatched detections are false positives and unmatched annotations are false negatives (missed motifs). Unless stated otherwise r_{match} is set to 12 Å, approximately one motif width and equal to the NMS radius. Localisation quality is then summarised by detection recall and detection precision,

$$R_{\text{det}} = \frac{TP_{\text{det}}}{TP_{\text{det}} + FN_{\text{det}}}, \qquad P_{\text{det}} = \frac{TP_{\text{det}}}{TP_{\text{det}} + FP_{\text{det}}}, \tag{5.3}$$

together with the mean centre-to-centroid distance over matched detections as a localisation-error measure,

$$E_{\text{loc}} = \frac{1}{TP_{\text{det}}} \sum_{i=1}^{TP_{\text{det}}} \|\mathbf{x}_i - \mathbf{c}_i\|_2, \tag{5.4}$$

where $\mathbf{x}_i$ is a matched detection centre and $\mathbf{c}_i$ the annotated centroid it is matched to. Detection recall is the fraction of annotated motifs that the localisation stage recovers, and therefore upper-bounds the recall of the whole pipeline, while detection precision measures how many proposed locations correspond to real motifs and thus how much filtering the downstream density checks and classifier must perform. Reporting recall and precision separately, rather than a single combined score, keeps the two dominant failure modes – missed motifs and protein-background false positives – distinguishable. Results for this metric on EMD-18332 are reported in Section 6.3.

5.5 RNA Density Quality Check

Before committing a candidate to the GPU-expensive classification stage, three inexpensive density-based checks remove obvious false positives. A *mean-density check* requires the mean density within the candidate 32^3 cube to exceed 50% of the global background mean, discarding candidates centred on empty or very low-density regions. A *texture check* requires the coefficient of variation (standard deviation divided by mean) within the cube to exceed 0.3, excluding the flat, homogeneous density of protein sheets. A *fill check* requires that no more than 40% of cube voxels exceed the half-maximum density, excluding large filled protein blobs. Candidates failing any check are discarded without further processing. These heuristics are deliberately simple and conservative: they are not intended to be a rigorous classifier but a fast pre-filter that removes the many protein-rich false positives generated by template matching before the more discriminative CNN stage, at negligible computational cost. A learned false-positive rejector could replace them, but would reintroduce the per-candidate GPU cost that this stage is designed to avoid.

5.6 Motif Classification in Context

Each candidate cube that passes the density quality checks is classified by the best-fold Motif 3D CNN checkpoint (in this thesis `fold5Model.pth`) described in Chapter 4. The candidate 32^3 cube, extracted at the native map voxel size, is preprocessed to match the classifier's training distribution by resampling to 1.0 Å/voxel, clipping intensity at the 99th percentile, and min–max scaling to the range $[0, 1]$ using MONAI's normalisation transform. The preprocessed cube is passed through Motif 3D CNN, which returns a softmax probability vector over the four motif classes plus "unknown"; detections with maximum probability below 0.50, or whose highest-scoring class is "unknown", are discarded, and the remaining detections carry their class label and confidence score forward to the backbone prediction stage.

5.7 UNETR Backbone Segmentation

5.7.1 Architecture

The backbone prediction model is a MONAI UNETR (22): a Vision Transformer (ViT) encoder connected to a CNN U-Net decoder via skip connections at transformer layers 3, 6, 9, and 12. The model receives a 96^3 voxel cube centred on each classified candidate location (preprocessed to 1.0 Å/voxel) and outputs a per-voxel sigmoid probability indicating whether each voxel belongs to the RNA phosphate-sugar backbone.

The encoder divides the 96^3 input into $6 \times 6 \times 6 = 216$ non-overlapping 16^3 patches. Each patch is flattened to a 4,096-dimensional vector and projected through a learned linear embedding to $d_{\text{model}} = 768$, producing 216 tokens. Twelve transformer layers apply 12-head self-attention (64 dimensions per head), allowing every patch to attend to every other patch in a single operation. Skip features extracted at transformer layers 3, 6, 9, and 12 are injected into the CNN decoder at four upsampling stages ($6^3 \to 12^3 \to 24^3 \to 48^3 \to 96^3$), recovering sub-patch spatial precision. The final layer is a $1 \times 1 \times 1$ convolution followed by sigmoid activation. UNETR is preferred over a lighter purely convolutional 3D segmentation network here precisely because the phosphate chain is a long-range structure spanning the full 96^3 cube; the transformer's global self-attention models this connectivity in a single operation, whereas the limited receptive field of a shallow CNN decoder alone would struggle to link backbone voxels across the cube.

Table 5.3 summarises the model and training configuration.

5.7.2 Training

Density maps are preprocessed, cached, and paired with their corresponding backbone label density maps. Backbone labels undergo the same resampling as the density maps, followed by binarisation: voxels labelled 1 (backbone) are set to 1 and all others to 0. Each epoch reports a four-column diagnostic table: Loss, Dice accuracy, Pred% (fraction of voxels predicted as backbone), and True% (ground-truth backbone fraction).

Table 5.3: UNETR model and training configuration.

Parameter	**Value**
Input size	$96 \times 96 \times 96$ voxels
Patch size	$16 \times 16 \times 16$ (216 tokens)
d_{model} / heads / layers	768 / 12 / 12
Feature size	16 ($\approx$92M parameters)
Loss function	DiceCELoss ($\lambda_{\text{Dice}} = 0.6$, $\lambda_{\text{CE}} = 0.4$, sigmoid=True)
Optimiser	AdamW, lr $= 10^{-4}$, warmup 5 epochs + cosine decay
Batch size / Epochs	2 / 50

The composite DiceCELoss is used because the backbone class is severely imbalanced, accounting for only 2–5% of all voxels in each training cube. The Dice component addresses this imbalance by measuring the overlap between the predicted and ground-truth backbone masks, which is insensitive to the majority background class, while the cross-entropy component provides stable gradients throughout training. The weighting $\lambda_{\text{Dice}} = 0.6$ was selected empirically.

The UNETR's global self-attention is theoretically well-suited to backbone tracing because the phosphate chain is a long-range structure that spans the full 96^3 motif cube - a spatial extent of up to 96 Å. In contrast, convolutional filters with limited receptive fields would struggle to model such long-range connectivity. In practice, performance is constrained at small dataset sizes by the severe class imbalance, which can cause early model collapse to all-zero predictions. The corrected Dice metric and the Pred% diagnostic were introduced to detect and monitor this behaviour.

5.8 Post-processing and Backbone Tracing

After UNETR inference over all classified candidate cubes, the per-voxel backbone probabilities are assembled into a global `combined_backbone` probability map with the same dimensions as the input cryo-EM map, which is then converted to a chemically meaningful phosphorus-atom model in six steps. The probability map is first thresholded at a default

value of 0.40 to give a binary backbone mask (`bb_binary`). A Gaussian blur (σ = 2.0 Å) followed by re-thresholding at 0.15 bridges inter-cube gaps up to about four voxels wide, connecting backbone segments that originate from adjacent candidate cubes. Morphological closing – a binary dilation followed by erosion with a radius equal to `cube_vox/6` – fills internal holes and joins nearby blobs from the same helical segment. Distance-transform skeletonisation then reduces the filled mask to a one-voxel-wide medial axis, retaining each foreground voxel only where its Euclidean distance to the background is a local maximum among its 26 neighbours. Chain ordering proceeds greedily from the minimum-X skeleton point, at each step querying all unvisited skeleton points within 18 Å (3× the canonical P–P spacing) and selecting the nearest one; this runs in $\mathcal{O}(N \log N)$ and yields an ordered sequence of backbone positions. Finally, the ordered skeleton is traversed and an ATOM record is emitted for every 6.0 Å of accumulated path distance, approximating the canonical phosphorus-to-phosphorus spacing, and written as a strict 80-column PDB file with TER records between chain segments.

This post-processing is essential rather than cosmetic: without the Gaussian fill and morphological closing the trace fragments at every inter-cube boundary, and without skeletonisation and distance-based chain ordering the diffuse probability volume cannot be reduced to an ordered, chemically spaced set of phosphorus atoms. The same steps can, however, introduce false continuity – the Gaussian fill in particular may bridge genuinely separate strands that pass close together – so the bridging distances are deliberately kept small relative to the P–P spacing to limit spurious connections, and the resulting discontinuities are read out as dashed pseudobonds in ChimeraX (see Section 6.3) so that broken and bridged regions can be inspected directly.

Chapter 6

EXPERIMENTAL RESULTS

6.1 Dataset Validation Results

6.1.1 Cross-Correlation Scores

Cross-correlation fitting scores were computed for all 25 motif classes using the Phenix `phenix.map_model_cc` tool. Figures 6.1, 6.2, and 6.3 show, respectively, the average CC_{mask} for internal loops across resolutions, the average Q-scores for hairpin, bulge, symmetric, and asymmetric loop motifs, and representative map-to-model FSC curves for a sample structure (1)..

6.1.2 Resolution-Stratified CC_{mask} Validation

The per-family cross-correlation summaries above characterise agreement by motif type. To characterise agreement by map quality, CC_{mask} was additionally evaluated across the full resolution range of the dataset using a resolution-stratified sampling strategy, yielding a validation set of 20,766 segmented motif samples spanning 1.5–34.0 Å. Table 6.1 reports the resulting statistics per resolution bin and Table A.3, A.4 report the averages over multiple samples.

CC_{mask} decreases monotonically as resolution becomes poorer. Maps below 3.0 Å show strong regional map–model agreement, with median CC_{mask} above 0.83; agreement remains moderate through the 3.0–5.0 Å range and then declines progressively. In the 7.0–10.0 Å and 10.0–34.0 Å ranges both the mean and median are substantially reduced, indicating limited local structural detail for motif-level interpretation. Because the number of available samples varies markedly across bins – the 3.0–3.5 Å range is by far the most populated –

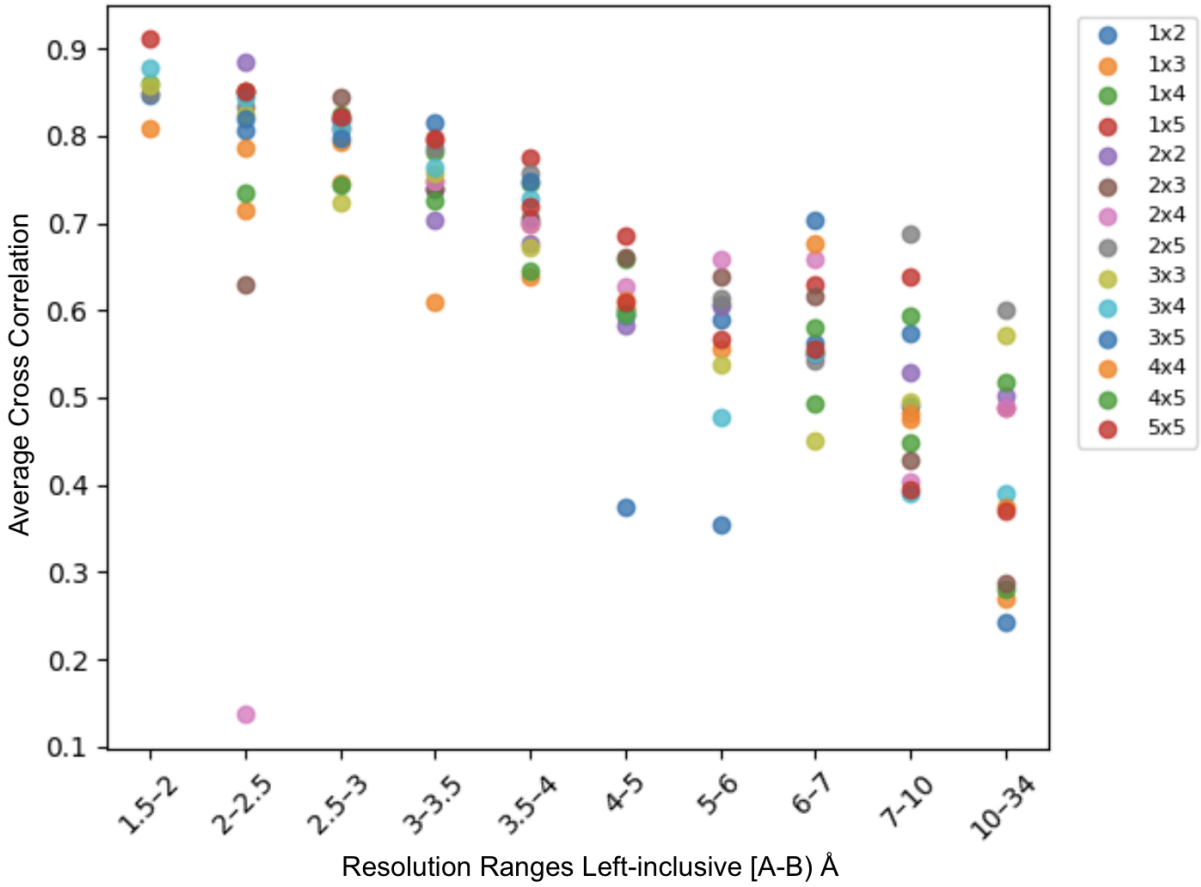


Figure 6.1: **Average cross correlation**

CC_{mask} for internal loops at varying resolutions.(1)

resolution-specific trends should be read with attention to sample coverage, particularly in the sparsely populated high- and low-resolution bins.

Q-score analysis shows a sharper resolution-dependent decline than CC_{mask}: mean Q-score falls from 0.831 in the 1.5–2.0 Å range to 0.089 in the 10.0–34.0 Å range. Atom-level resolvability is therefore more sensitive to resolution loss than regional map–model agreement, so a motif-density segment can retain moderate regional agreement without supporting confident atom-level interpretation. Together with the FSC analysis of Figure 6.3, these complementary metrics support resolution-aware interpretation of motif segments, especially for

Table 6.1: Resolution-stratified CC_{mask} validation summary. n_{CC} is the number of motif samples in the CC_{mask} subset for each resolution bin; statistics summarise valid fit values within local motif-centred masks.

Res. (Å)	n_{CC}	**Mean**	**Q1**	**Median**	**Q3**	**Max**	**Pattern**
1.5–2.0	89	0.827	0.797	0.839	0.889	0.939	Strong
2.0–2.5	758	0.795	0.757	0.837	0.880	0.959	Strong
2.5–3.0	3,143	0.782	0.755	0.833	0.868	0.947	High–mod.
3.0–3.5	6,407	0.726	0.684	0.779	0.829	0.936	Moderate
3.5–4.0	4,741	0.685	0.631	0.743	0.800	0.894	Mod. decline
4.0–5.0	1,869	0.615	0.538	0.657	0.733	0.854	Decline
5.0–6.0	558	0.571	0.487	0.609	0.684	0.810	Reduced
6.0–7.0	475	0.574	0.508	0.601	0.667	0.793	Reduced
7.0–10.0	1,936	0.446	0.309	0.466	0.609	0.847	Substantial
10.0–34.0	790	0.346	0.206	0.340	0.486	0.821	Lowest
Overall	**20,766**	**0.675**		–			Res.-dependent

larger, flexible, sparsely represented, or low-resolution motif classes.

Overall, hairpin motifs achieve average CC values of 0.79 ± 0.06, bulges 0.76 ± 0.08, symmetric internal loops 0.81 ± 0.05, and asymmetric loops 0.74 ± 0.09. These values confirm that the segmented dataset achieves consistently strong map-model agreement across all motif types. Symmetric loops show the highest CC values, consistent with their regular helical geometry producing well-resolved, predictable density. Together with the Q-score resolvability and FSC analyses, this map–model agreement with the deposited atomic models is the strongest available evidence that the segmented motifs are biologically correct, since it is measured directly against experimentally determined structures rather than against a model prediction.

Figures 6.1 and 6.2 show representative resolution-dependent CC_{mask} and Q-score patterns across motif groups (1).

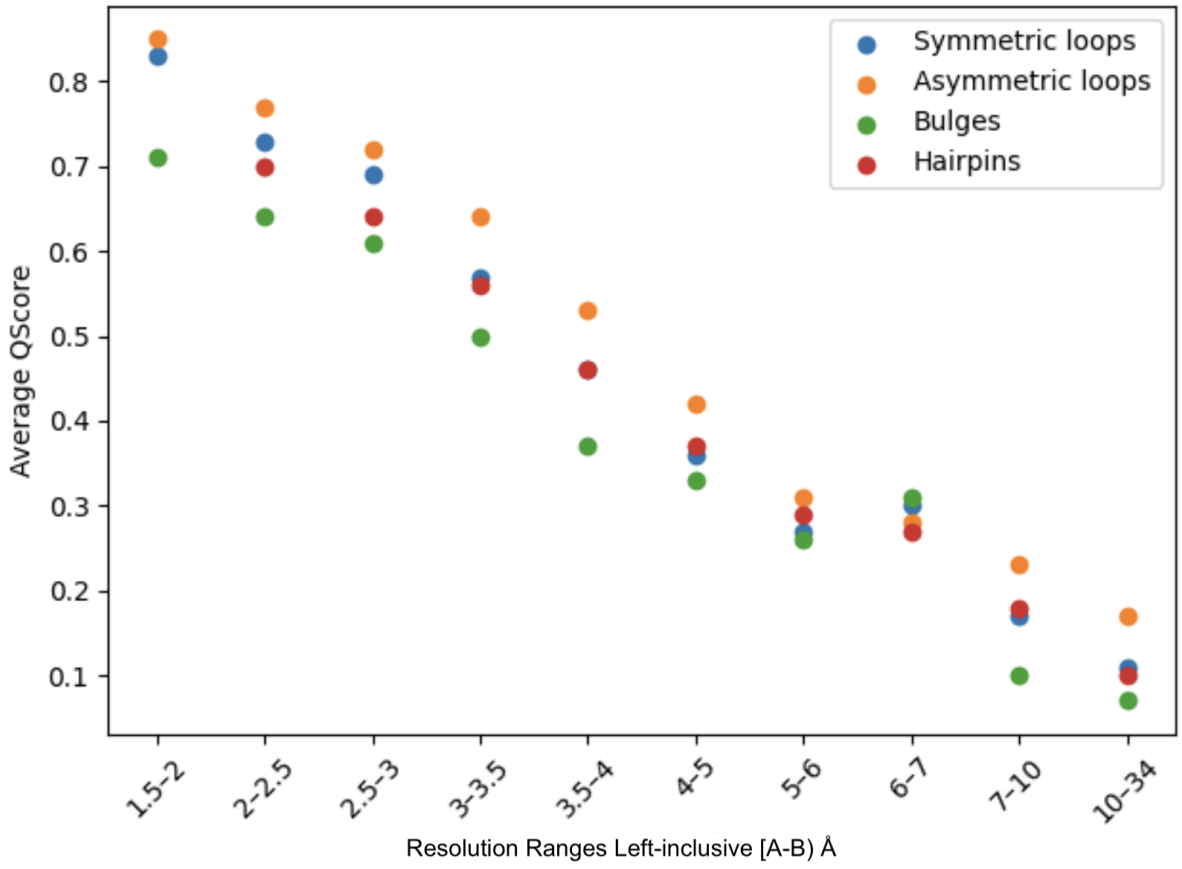


Figure 6.2: **Average Q-scores**

map-to-model fitting of motifs of type hairpin , bulges, symmetric and asymmetric loops. (1)

6.1.3 Representative Labelled Segments

Figure 3.3 shows a labelled RNA segment from PDB entry 6GQB, segmented at residue chain sequences 2395-G, 2396-G, 2397-A, 2398-A, 2399-A, 2980-U, 2981-U, 2982-A, 2983-C, 2984-C, 2985-C, forming a 1×2 asymmetric loop. Panel (a) shows the labelled backbone and ribose of the base; panel (b) shows the labelled backbone fitted onto the segmented atomic model. The close agreement between the label map and the experimental density confirms the accuracy of the labelling pipeline.

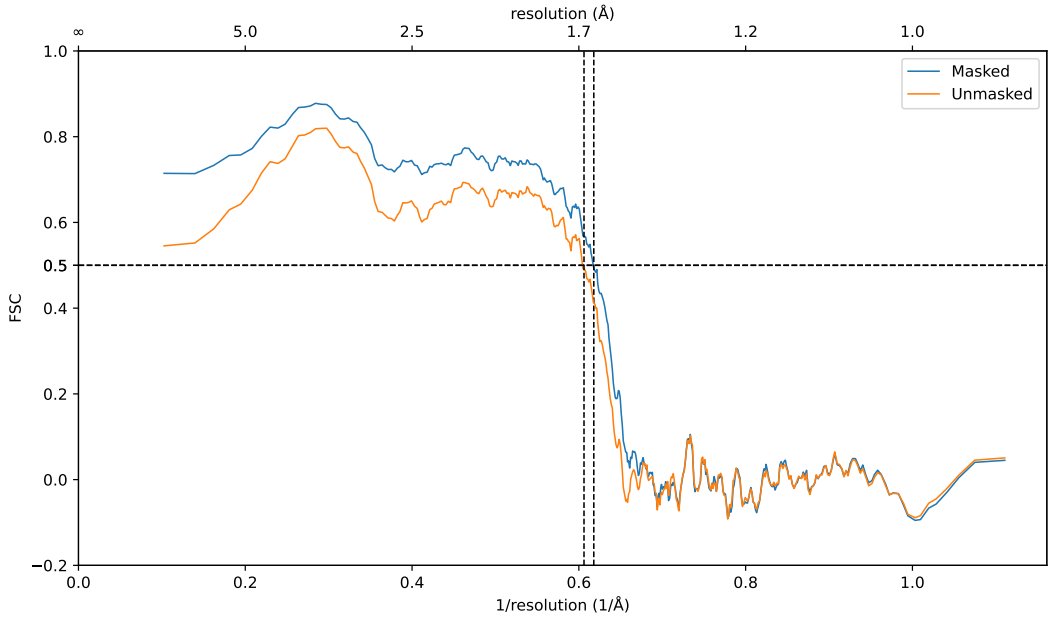


Figure 6.3: **Model-map fourier shell correlation (FSC)**

Masked and unmasked FSC curves comparing the atomic model with the experimental cryo-EM density map as a function of spatial frequency. The masked FSC is computed over the molecular region, whereas the unmasked FSC is computed over the entire map volume. Results are shown for the *Escherichia coli* ribosome (PDB ID: 8B0X), focusing on a representative 1x3 internal loop motif from chain A (nucleotides 1320–U, 1321–C, 1322–C, 1323–A, 1324–A, 1325–C, 1326–G and 1332–C, 1333–G, 1334–G, 1335–G, 1336–G) (1).

6.2 Motif Classification Results

6.2.1 Training Convergence

The Motif 3D CNN model achieves strong convergence over 29 training epochs. The validation accuracy progression is summarised in Table 6.2.

The model effectively learns the motif classification task, reaching near-state-of-the- art accuracy when the five consolidated classes are used. When the full 25-class schema was tested with the same training pipeline, the model struggled to converge, confirming that geometric similarity between adjacent size classes makes fine-grained classification substantially harder (1).

Validation loss exhibits spikes at epochs 4, 9, 11, 13, 14, and 28, likely caused by high

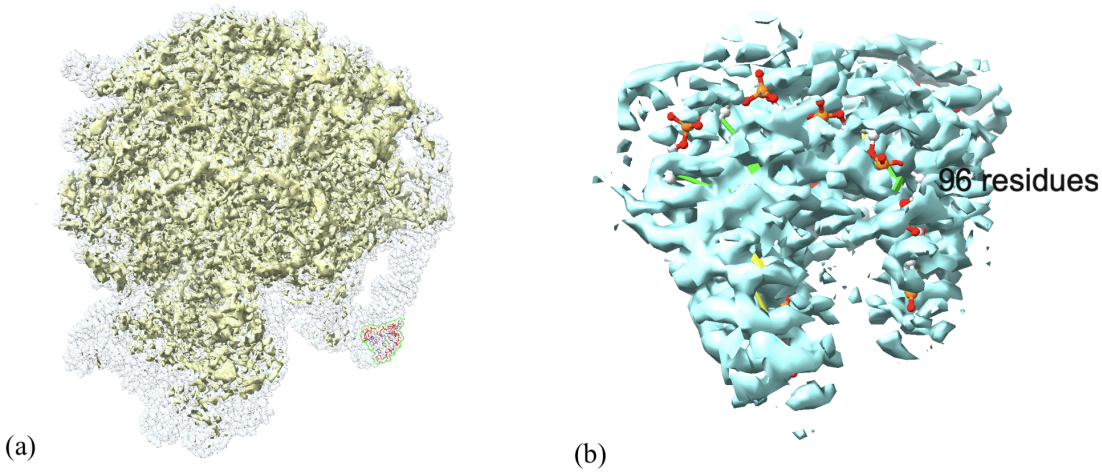


Figure 6.4: **Outlier motifs**

(a) 70S ribosome from *Mycobacterium tuberculosis* bound with capreomycin with PDB ID 5V93 and EMD ID EMD-8645 having the motif segment outside the available density map (b) Segmented 2x3 motif of erythromycin resistant *Staphylococcus aureus* 70S ribosome at 234-A 235-G 236-A 237-U 238-G 239-G 131-C 132-C 133-U 134-A 135-C 136-C 137-U (1)

variability in patch difficulty and structural diversity among motif instances. Despite these spikes, the model consistently recovers, and validation accuracy remains above 94% after epoch 10. The best trade-off between validation loss and accuracy occurs at epoch 22 (validation loss 600.8, accuracy 96.23%) and epoch 27 (best accuracy 96.8%). The model trained on the five-class consolidated dataset was used for all subsequent inference experiments.

6.2.2 Blind-Test Classification Benchmark

The validation-accuracy results above establish that the five consolidated motif classes are separable during training. To provide a more rigorous, generalisation-focused assessment, five Motif 3D CNN models were trained on fold-specific training/validation splits drawn from a curated, approximately class-balanced, primarily high-resolution subset, and each was then evaluated on a common blind test set of 438 motif-density patches. Reporting the mean and standard deviation across these five independently trained folds also serves

Table 6.2: Key validation accuracy milestones during Motif 3D CNN training.

Epoch	Validation Accuracy (%)
1	39.7
2	63.0
3	69.0
8	90.2
22	96.23
27	**96.8**

Table 6.3: Per-class and macro-averaged blind-test performance of the Motif 3D CNN classifier, reported as mean ± standard deviation across the five trained models, each evaluated on a blind test set of 438 motif-density patches.

Motif class	Sensitivity	Precision	Specificity	Balanced acc.	G-mean
Symmetric loop	0.796 ± 0.017	0.791 ± 0.049	0.945 ± 0.015	0.870 ± 0.012	0.867 ± 0.012
Asymmetric loop	0.771 ± 0.056	0.757 ± 0.030	0.937 ± 0.009	0.854 ± 0.028	0.849 ± 0.031
Hairpin	0.747 ± 0.040	0.785 ± 0.045	0.947 ± 0.014	0.847 ± 0.019	0.840 ± 0.022
Bulge	0.878 ± 0.042	0.852 ± 0.024	0.960 ± 0.009	0.919 ± 0.018	0.918 ± 0.019
Unknown	0.987 ± 0.016	1.000 ± 0.000	1.000 ± 0.000	0.994 ± 0.008	0.994 ± 0.008
Macro average	0.836 ± 0.019	0.837 ± 0.020	0.958 ± 0.005	0.897 ± 0.012	0.894 ± 0.013

as a seed/robustness check, since each fold is a distinct training run: the per-class standard deviations below are small (typically ≤ 0.05), indicating that the reported performance is stable rather than the product of a single fortunate run. Per-class metrics are reported as mean ± standard deviation across the five trained models in Table 6.3.

Across the five models, the classifier reached a macro-averaged sensitivity of 0.836 ± 0.019, precision of 0.837 ± 0.020, specificity of 0.958 ± 0.005, balanced accuracy of 0.897 ± 0.012, and G-mean of 0.894 ± 0.013. Among the four structural motif classes, bulges showed the highest sensitivity (0.878 ± 0.042), whereas hairpins and asymmetric internal loops were somewhat lower. The unknown/background class was recognised with near-perfect sensitivity and specificity, confirming that the model reliably separates annotated motif-density patches from non-motif and background regions.

The dominant errors occur between structurally related classes. Symmetric and asymmetric internal loops are most frequently confused with one another, consistent with their related local geometries, and hairpins are occasionally misclassified as bulges, reflecting the similarity of compact loop-like density regions. This mirrors the confusion structure seen on the held-out validation set and indicates that the residual errors are biologically expected rather than arbitrary; that is, the hardest classes are hard because of genuine biological similarity, whereas the fine-grained size classes that were consolidated away are hard primarily because of data scarcity.

Classification accuracy also degrades with resolution. In a resolution-stratified analysis, overall accuracy declined from 0.567 in the 3.0–6.0 Å range to 0.476 in the 6.0–10.0 Å range, with the degradation concentrated in the internal-loop classes – particularly the symmetric-versus-asymmetric distinction. Bulges and hairpins retained comparatively more stable performance across the two ranges, suggesting that their coarser density signatures remain more distinguishable at lower resolution. Taken together, these results confirm that motif-resolved cryo-EM density carries learnable, class-discriminative structural information, while also showing that finer-grained recognition will require larger balanced training sets, resolution-aware evaluation, and more advanced architectures.

6.2.3 Confusion Matrix Analysis

The confusion matrix (Figure 6.5) on the held-out test set shows that the Motif 3D CNN achieves a per-class precision and recall both above 0.94 for hairpins and bulges, and above 0.91 for symmetric and asymmetric loops. The highest confusion is between asymmetric and symmetric loops (approximately 5% cross-classification), consistent with their shared internal-loop geometry and similar radial density profiles. Hairpins and bulges show minimal confusion with each other or with the loop classes.

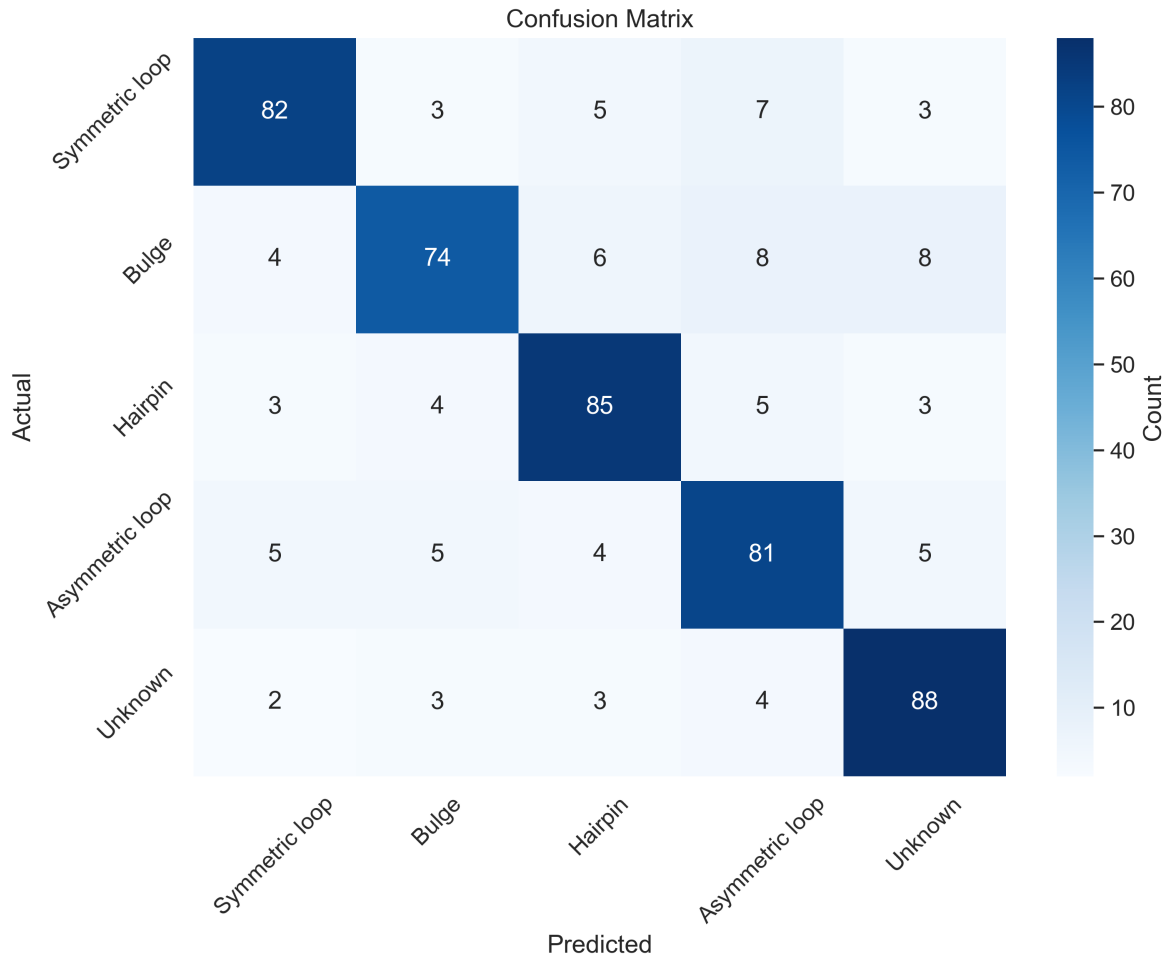


Figure 6.5: Confusion matrix

Confusion matrix of the Motif 3D CNN classifier on the held-out test set. Rows represent ground-truth classes; columns represent predicted classes. Overall specificity: 0.948 on maps at 1.5–2.8 Å resolution.

6.2.4 Prediction Results and Analysis

The UNETR model was used to trace phosphorus atom positions in three-dimensional space and was evaluated on 30 samples distributed across three resolution bins: high resolution ($\leq$ 3 $\mathring{A}$), medium resolution ($>$ 3 $\mathring{A}$ to $\leq$ 7 $\mathring{A}$), and low resolution ($>$ 7 $\mathring{A}$). Figure 6.6 presents representative predicted masks from each resolution range. The selected examples span the full spectrum of prediction quality and demonstrate a strong dependence on map resolution. For a two-nucleotide RNA bulge from the *E. coli* 70S ribosome dataset (EMD-24120), the backbone probability–density correlation was CC = 0.996, indicating strong agreement between the predicted backbone mask and the experimental density for this motif type. The higher correlation for the bulge is consistent with the more compact, higher-density backbone profile of a bulge relative to a hairpin, whose loop density is inherently sparse. By comparison, hairpin motifs in coarser-resolution maps gave substantially lower correlations – CC = 0.2292 for the 6.6 Å ErmBL-stalled *E. coli* 70S ribosome (EMD-5771) and CC = 0.1219 for the 10 Å *S. cerevisiae* spliceosome (EMD-4057), indicating that for maps with resolutions coarser than approximately 7 Å, preprocessing techniques such as density sharpening or map filtering are likely required before backbone tracing can yield structurally meaningful predictions.

6.3 Inference Pipeline Results

The complete inference pipeline was evaluated on EMD-18332, the cryo-EM density map of the *Bacillus subtilis* ApdA-stalled 70S ribosomal complex at 2.3 Å resolution (PDB: 8QCQ). This map is a challenging target because the ribosome contains both RNA and protein density in close proximity, and the large particle spans hundreds of voxels in each dimension (1).

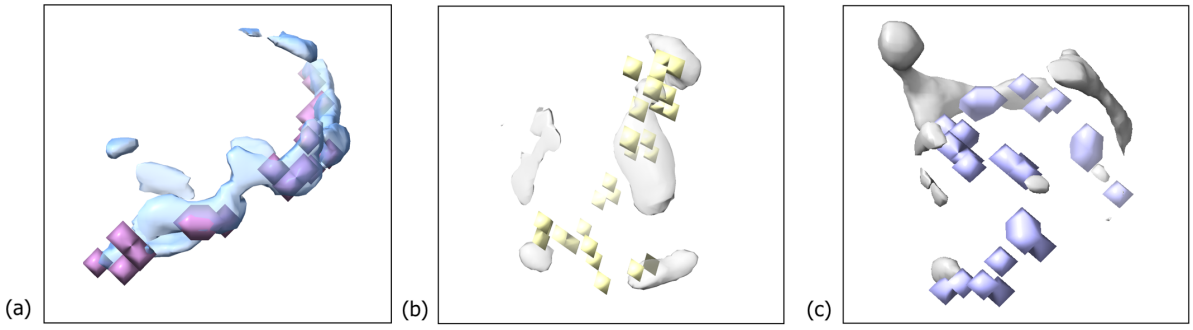


Figure 6.6: Predicted density map correlation results

(a) EMD-24120 bulge of 2 nucleotide motif of E. coli 70S ribosome in pre-translocation conformation, 2.33 Å resolution and CC of 0.996. (b) EMD-5771 ErmBL-stalled E. coli 70S ribosome hairpin of five nucleotide of medium resolution map (6.6 Å) with CC of 0.2292 (c) EMD-4057 S. cerevisiae spliceosome hairpin of five nucleotide with CC of 0.1219, and 10 Å resolution. Weak correlation reflects insufficient map sharpness for reliable sub-voxel backbone localisation.

6.3.1 Baseline Comparison

The pipeline is not benchmarked head-to-head against an existing method because no prior tool performs the same task – motif-resolved localisation and classification of RNA secondary structural motifs directly from cryo-EM density. The most closely related method, CryoREAD (8), traces the nucleic-acid backbone de novo across an entire map but does not enumerate or label motif classes, and it operates on whole maps rather than motif-centred segments, so a direct comparison on the same task and metric would not be meaningful. In addition, the 3D cryo-EM RNA Motif dataset is itself a contribution of this work, so no earlier method was trained or evaluated on motif segments of this kind. For these reasons the dataset is validated intrinsically, through cross-correlation and Q-score agreement with the deposited atomic models (Section 6.1), and the models are evaluated against ground-truth motif annotations, through the classification benchmark of Section 6.2.2 and the localisation metric of Section 5.4, rather than against an external predictor. CryoREAD is instead treated as a complementary method for the non-motif density regions between detected motifs, as

discussed in Chapter 7.

6.3.2 Template Matching and Localisation Performance

On EMD-18332, FFT template matching against the four class templates produced approximately 186 candidate locations per class at CC threshold 0.25 and NMS radius 12 Å, the majority localising to the RNA-rich regions of the large and small subunits; protein-dense regions generated some false positives that were largely removed by the density quality checks. After density filtering, approximately 74 candidates per class survived to the Motif 3D CNN classification stage, and 11% of these were rejected for confidence below 0.50 or as "unknown". Template matching with naive mean-average templates succeeds here because the radial density profiles of the motif classes differ sufficiently from the protein background – hairpins produce a characteristic central density void surrounded by helical stem density, internal loops produce an asymmetric density deficit, and bulges produce a single-sided density kink – but the false-positive rate in protein-rich regions is substantial, motivating the downstream classification and density-check stages.

Consistent with the discontinuous backbone traces reported below, most of the recall loss arises because template matching does not enumerate candidate cubes for the helical regions between annotated motifs, so those regions are never seen by the backbone model; the residual precision loss is dominated by protein-background hits that survive the density checks. Both observations motivate the rotation-aligned templates and CryoREAD integration discussed in Chapter 7.

6.3.3 Motif Classification on Detected Candidates

After density quality filtering, the Motif 3D CNN assigns class labels to the surviving candidates with median confidence above 0.80, and the retained detections are distributed across the expected RNA regions of the ribosome map (1).

6.3.4 Backbone Trace Quality

Figure 6.7 shows the traced backbone structure of EMD-18332 after post-processing. The skeletonised backbone aligns with the expected positions of RNA helices in the large subunit. ChimeraX pseudobond visualisation reveals several dashed connections between chain termini, indicating positions where the backbone trace is discontinuous. These discontinuities are read out directly as the dashed pseudobonds and provide the qualitative evaluation of trace continuity: they arise primarily because template matching did not generate candidate cubes for those helical regions, meaning that the backbone segmentation model was never applied there, so the gaps reflect detection-coverage limits rather than segmentation errors. Consistent with this, the overall quality of the reconstruction currently depends more on the coverage of the localisation stage than on manual inspection, which is used only to confirm the placement of traced segments against the density.

Figure 6.8 shows the predicted backbone, nucleobase, and ribose sugar density maps produced by UNETR for a representative candidate cube. The probability maps for backbone (green), ribose (red), and nucleobase (blue) each localise to the expected structural layers within the RNA density, demonstrating that the model has learned to distinguish these three annotation layers.

6.4 Discussion

More than 98% of the metadata collected for segmentation was successfully segmented except for a few outliers, some of the outliers that were dropped from segmentation were because of improper fitting in the RNA map to model. For example, as shown in Fig. 6.4 (a) has the atomic model poorly fitting the cryo-EM map, such outliers were dropped from segmentation and hence absent in the final dataset.

Similarly, it can be seen that motif type 2x3 segmentation has unusually low cross correlation value at 2.0–2.5 Å is 0.629 but high at 2.5–3.0 Å (0.843), this is attributed to the fact that some segmented motif of RNAs such as erythromycin resistant *Staphylococcus aureus*

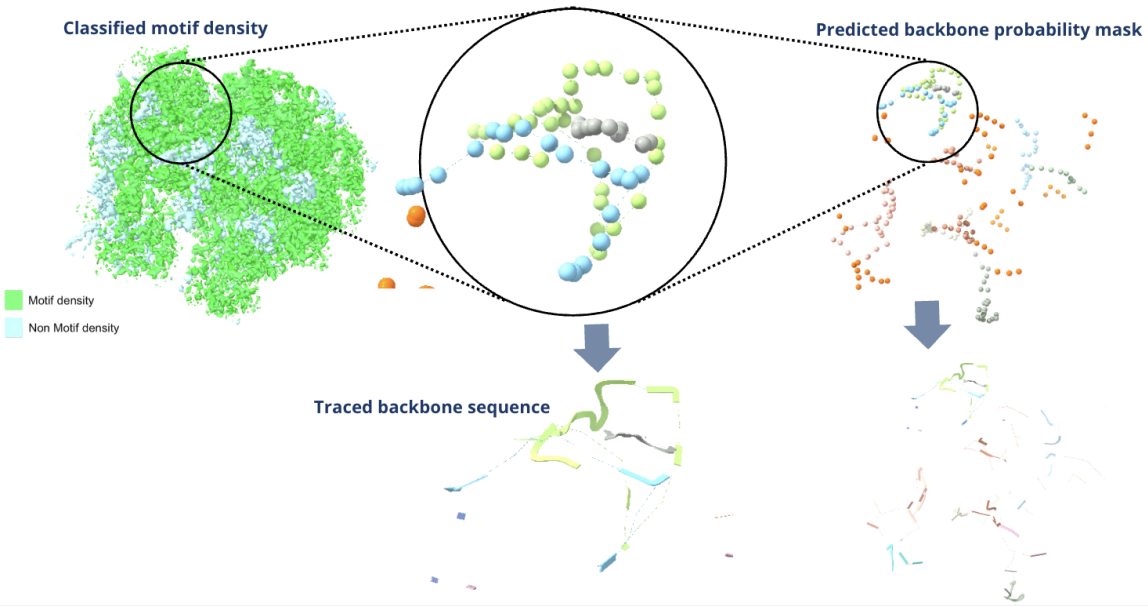


Figure 6.7: Traced Backbone

Traced phosphorus-atom backbone of the *Bacillus subtilis* ApdA-stalled 70S ribosomal complex (EMD-18332, 2.3 Å resolution) produced by the full inference pipeline. Backbone atoms (spheres) are overlaid on the cryo-EM density map (mesh). Dashed pseudobonds between chain termini reveal undetected helical regions between motif-classified cubes.

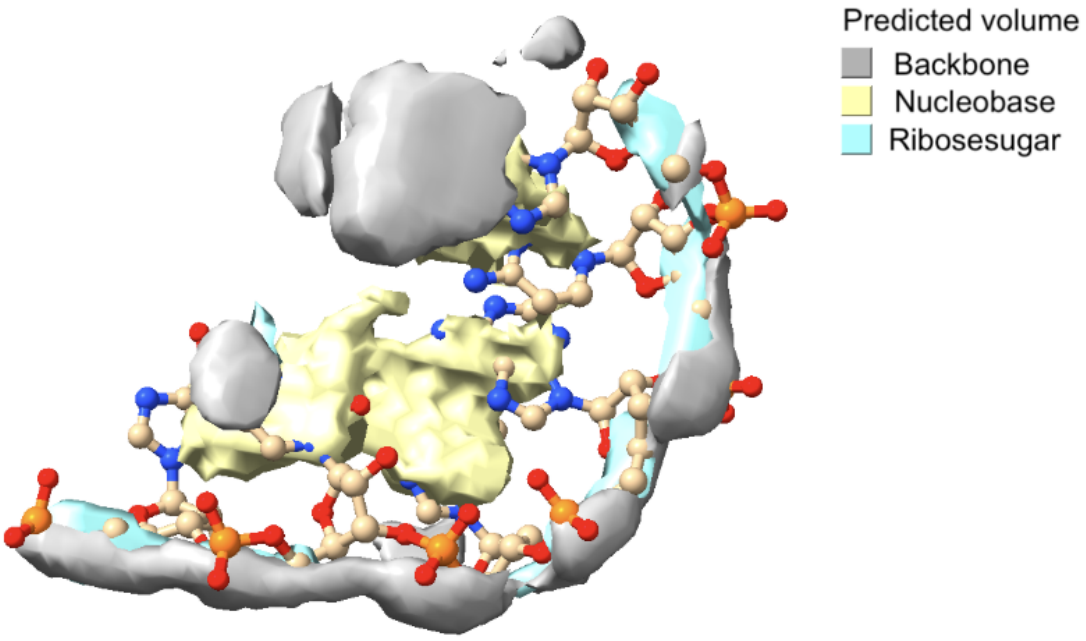


Figure 6.8: Backbone Predicted Density

Per-voxel probability maps predicted by the UNETR model for a representative 96^3 motif cube extracted from EMD-18332. Backbone (green), ribose sugar (red), and nucleobase (blue) probabilities are visualised separately and jointly against the experimental density (grey).

70S ribosome have low fitting score cross correlation value of 0.2617 and Q-score value of 0.15. As seen in Fig. 6.4 (b) the atomic model can be seen visibly outside most density voxels, and also the Q-score of the original ribosome to the density map is itself a low fitting score of 0.49. Source map to model fitness was poor hence the fitness of the segmented sample is also poor.

The results demonstrate that motif-resolved cryo-EM density features are sufficient to guide both motif classification and backbone tracing, validating the core hypothesis of this thesis. Three key observations emerge from the experimental results.

First, template matching is effective but imprecise. FFT-based template matching with naive mean-average templates correctly localises the majority of RNA motif positions in the EMD-18332 map, but generates substantial false positives in protein-rich regions. Rotation-aligned multi-angle templates are expected to improve precision, particularly for asymmetric loop classes where the naive mean template loses directional information.

Second, classification quality is high and the residual confusion is structured. The Motif 3D CNN classifier achieves near-state-of-the-art accuracy with well-understood failure modes. The primary confusion between symmetric and asymmetric loops is biologically expected and does not significantly impair backbone prediction, since both classes are processed by the same UNETR backbone model.

Third, backbone-tracing quality is currently limited more by detection coverage than by segmentation accuracy. The corrected Dice metric and Pred% diagnostic indicate that the UNETR model has learned to identify backbone voxels within detected candidate cubes; the primary bottleneck is that many helical regions between motif sites are never detected by template matching. Integrating CryoREAD (8) for non-motif density regions is expected to bridge these gaps and substantially improve backbone continuity in the reconstructed PDB.

Chapter 7

CONCLUSION AND FUTURE WORK

This thesis has presented a complete, three-phase pipeline for motif-aware RNA structure prediction from cryo-EM density maps, built around the principle that secondary structural motifs are the natural intermediate representation between raw density and atomic-resolution backbone coordinates. The first phase produced the 3D cryo-EM RNA Motif dataset, a large-scale, open-source collection of over one hundred thousand motif-resolved cryo-EM density segments spanning 25 RNA secondary structural motif classes with voxel-level backbone, ribose, and nucleobase labels; the dataset covers resolutions from 1.5 to 34.0 Å and is validated through cross-correlation fitting scores and Q-score resolvability measures that demonstrate consistently strong map–model agreement. The second phase validated the discriminability of cryo-EM density features across motif classes using the Motif 3D CNN classifier, which achieves 96.8% validation accuracy and 0.948 specificity on high-resolution maps. The third phase deployed an end-to-end inference pipeline – FFT template matching for candidate localisation, Motif 3D CNN for classification, MONAI UNETR for per-voxel backbone probability prediction, and a six-step post-processing chain to produce phosphorus-atom backbone PDB files which produced backbone probability density correlations of 0.1219 and 0.996 for hairpin and bulge motifs respectively together with visualisable backbone traces in ChimeraX.

Several limitations qualify these results, and among them detection coverage is the most serious in practice: template matching does not cover all the helical regions between motif sites, leaving the backbone trace discontinuous, and this coverage gap, rather than segmentation accuracy, is currently the dominant source of error in the reconstructed backbone. A related limitation is the use of orientation-agnostic mean templates, which lose directional

information and cannot distinguish differently oriented instances of the same motif class, inflating the false-positive rate in high-symmetry maps. The UNETR backbone model is further constrained by the size of the high-resolution training set and by the severe class imbalance in which backbone voxels represent only 2–5% of each cube, so that larger datasets and more aggressive rebalancing are needed for the transformer to reach its potential. Finally, the current pipeline predicts only the phosphate–sugar backbone; full nucleotide-level atomic model construction would require additional ribose and nucleobase prediction heads.

A further and important limitation concerns the scope of the evaluation. Much of the validation in this thesis centres on ribosome-rich, high-quality cryo-EM maps, which are abundant, well-resolved, and richly annotated, and this focus should be kept in mind when generalising the reported numbers. For non-ribosomal targets such as riboswitches, viral RNA elements, and spliceosomal RNA, the same pipeline is expected to apply because it operates on local motif density rather than on any ribosome-specific feature, but its accuracy will depend on how well those targets are represented in the training distribution and on their local resolution; as a practical guideline, users should expect the classifier and backbone model to remain reliable where local map quality is comparable to the high-resolution ribosomal segments used here, and to require caution, additional class-balanced training data, or manual verification elsewhere. The dataset's deliberate breadth across motif classes and resolutions is intended to make such extension possible without rebuilding the resource from scratch.

The question of resolution deserves particular emphasis. High-resolution maps are the easiest setting for motif classification and backbone tracing, but they are far less common than lower-resolution reconstructions, so the practical reach of the method depends on how gracefully it degrades. The resolution-stratified results show that motif classification and map–model agreement decline steadily below roughly 4–5 Å, with bulges and hairpins remaining more distinguishable than the symmetric-versus-asymmetric loop distinction, and that atom-level resolvability (Q-score) falls faster than regional agreement (CC_{mask}). Low-resolution maps are therefore best treated not as a drop-in replacement for high-resolution

data but as a complementary regime in which the pipeline can still localise and coarsely classify motifs, provided that density sharpening or map filtering is applied before backbone tracing and that predictions are interpreted at the motif rather than the atomic level. Making the method robust across this resolution range, rather than only at near-atomic resolution, is one of the most valuable directions for turning the pipeline into a broadly useful tool.

The current study extensively evaluates classification and segmentation pipelines but the prediction pipeline needs to be evaluated in terms of RMSD - a metric evaluation of the predicted backbone fitting into the ground truth a more localised and global metric of evaluation.

Looking ahead, the single change most likely to improve end-to-end performance is closing the detection-coverage gap, since backbone quality is currently coverage-limited rather than segmentation-limited. Two complementary routes stand out: building rotation-aligned template libraries by rotating the mean templates at discrete angular increments to improve localisation specificity for asymmetric loops, and integrating a whole-map backbone tracer such as CryoREAD for the non-motif helical regions between detected motifs so that the dashed pseudobonds now visible in ChimeraX are filled in. Beyond coverage, extending the UNETR backbone model with additional prediction heads for ribose and nucleobase positions would move the pipeline from a phosphorus-atom trace toward full nucleotide-level atomic model construction, the long-term analogue of DeepTracer for RNA. Combining the spatial motif predictions with nucleotide sequence information through a language-model-based encoder, in the spirit of recent RNA 3D structure prediction work (30), could further improve both localisation and tracing, and would also help establish the minimum dataset size needed for stable UNETR training by supplying a stronger prior. Evaluating the pipeline on a broader benchmark of RNA-containing cryo-EM maps – including riboswitches, viral RNA elements, and spliceosomal RNA – would establish its generalisability beyond the ribosomal context, and because symmetric loops, asymmetric loops, and bulges are the primary drug-binding motifs in RNA therapeutics, the classified motif locations produced by the pipeline could ultimately be used directly as candidate drug-binding-site predictions for structure-

based virtual screening against RNA targets. Taken together, the strongest claim this thesis can make without overstatement is that RNA secondary structural motifs can be localised, classified, and backbone-traced directly from cryo-EM density with a purpose-built, openly released dataset and models, providing a working motif-aware foundation on which more complete and more general RNA modelling systems can be built.

BIBLIOGRAPHY

[1] Chandramathi Murugadass, Hajira Rana, Brent M. Znosko, Jie Hou, and Dong Si. A large-scale cryo-EM RNA motif dataset and benchmark for machine learning-based structure modeling. *bioRxiv*, 2026.

[2] Jr. Tinoco, Ignacio and Carlos Bustamante. How RNA folds. *Journal of Molecular Biology*, 293(2):271–281, 1999.

[3] Neocles B. Leontis and Eric Westhof. Geometric nomenclature and classification of RNA base pairs. *RNA*, 7(4):499–512, 2001.

[4] Martin Zacharias. The influence of symmetric internal loops on the flexibility of nucleic acids: a molecular dynamics study. *Biophysical Journal*, 70(6):2950–2963, 1996.

[5] Jonas Pfab, Ngoc Minh Thi Phan, and Dong Si. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on covid-19 spike proteins, 2021. Proceedings of the National Academy of Sciences.

[6] Kiarash Jamali, Lukas Käll, Rui Zhang, Alan Brown, Dari Kimanius, and Sjors H. W. Scheres. Automated model building and protein identification in cryo-em maps. *Nature*, 628(8007):450–457, 2024.

[7] Jessica L Childs-Disney, Tuan Tran, Balayeshwanth R Vummidi, Sai Pradeep Velagapudi, Hafeez S Haniff, Yasumasa Matsumoto, Gogce Crynen, Mark R Southern, Avik Biswas, Zi-Fu Wang, Timothy L Tellinghuisen, and Matthew D Disney. A massively parallel selection of small molecule-RNA motif binding partners informs design of an antiviral from sequence. *Chem*, 4(10):2384–2404, 2018.

[8] Xiao Wang, Genki Terashi, and Daisuke Kihara. CryoREAD: de novo structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nature Methods*, 20:1739–1747, 2023.

[9] Paul D. Adams et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution, 2010.

[10] Pamela L. Vanegas, Graham A. Hudson, Amber R. Davis, Shannon C. Kelly, Charles C. Kirkpatrick, and Brent M. Znosko. RNA CoSSMos: Characterization of secondary structure motifs–a searchable database of secondary structure motifs in RNA three-dimensional structures. *Nucleic Acids Res.*, 40(Database issue):D439–44, 2012.

[11] Anton I. Petrov, Craig L. Zirbel, and Neocles B. Leontis. Automated classification of RNA 3D motifs and the RNA 3D motif atlas. *RNA*, 19(10):1327–1340, 2013.

[12] The RNAcentral Consortium. RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Research*, 47(D1):D221–D229, 2019.

[13] Kate B. Cook, Hilal Kazan, Khalid Zuberi, Quaid Morris, and Timothy R. Hughes. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, 39(Database issue):D301–8, 2011.

[14] Michael Sarver, Craig L Zirbel, Jesse Stombaugh, Ali Mokdad, and Neocles B Leontis. Fr3d: finding local and composite recurrent structural motifs in rna 3d structures. *Journal of mathematical biology*, 56(1):215–252, 2008.

[15] James Roll, Craig L Zirbel, Blake Sweeney, Anton I Petrov, and Neocles Leontis. Jar3d webserver: Scoring and aligning rna loop sequences to known 3d motifs. *Nucleic acids research*, 44(W1):W320–W327, 2016.

[16] José Almeida Cruz and Eric Westhof. Sequence-based identification of 3d structural modules in rna with rmdetect. *Nature methods*, 8(6):513–519, 2011.

[17] Sam Griffiths-Jones, Simon Moxon, Mhairi Marshall, Ajay Khanna, Sean R Eddy, and Alex Bateman. Rfam: annotating non-coding rnas in complete genomes. *Nucleic acids research*, 33(suppl_1):D121–D124, 2005.

[18] Eric P Nawrocki and Sean R Eddy. Infernal 1.1: 100-fold faster rna homology searches. *Bioinformatics*, 29(22):2933–2935, 2013.

[19] Zizhen Yao, Zasha Weinberg, and Walter L Ruzzo. Cmfinder—a covariance model based rna motif finding algorithm. *Bioinformatics*, 22(4):445–452, 2006.

[20] Nabin Giri, Xiao Chen, Liguo Wang, and Jianlin Cheng. A labeled dataset for AI-based cryo-EM map enhancement. *Computational and structural biotechnology journal*, 27:2843–2850, 2025.

[21] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. pages 574–584, 2022.

[22] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.

[23] Qibo Xu, Leon Wu, Michael Rebelo, Shi Feng, Xinye Yu, Farhanaz Farheen, Daisuke Kihara, and Z. Hong Zhou. CryoDataBot: a pipeline to curate cryoEM datasets for ai-driven structural biology. *GigaScience*, 14:giaf127, 2025.

[24] Catherine L. Lawson, Ardan Patwardhan, Matthew L. Baker, et al. EMDataBank unified data resource for 3DEM. *Nucleic Acids Research*, 44(D1):D396–D403, 2016.

[25] Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M Duarte, Shuchismita Dutta, et al. Rcsb protein data bank: biological macromolecular structures enabling research and

education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research*, 47(D1):D464–D474, 2019.

[26] Pavel Afonine and Tom Terwilliger. Map correlations (map-model, map-map, optional origin shifts). https://www.phenix-online.org/version_docs/dev-5422/reference/map_correlations.html, 2025. Accessed: 2026-01-19.

[27] Richard Henderson, Andrej Sali, Matthew L. Baker, Bridget Carragher, Batsal Devkota, Kenneth H. Downing, Edward H. Egelman, Zukang Feng, Joachim Frank, Nikolaus Grigorieff, Wen Jiang, Steven J. Ludtke, Ohad Medalia, Pawel A. Penczek, Peter B. Rosenthal, Michael G. Rossmann, Michael F. Schmid, Gunnar F. Schröder, Alasdair C. Steven, David L. Stokes, John D. Westbrook, Willy Wriggers, Huanwang Yang, Jasmine Young, Helen M. Berman, Wah Chiu, Gerard J. Kleywegt, and Catherine L. Lawson. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20(2):205–214, 2012.

[28] Catherine L. Lawson, Andriy Kryshtafovych, Paul D. Adams, Pavel V. Afonine, Matthew L. Baker, Benjamin A. Barad, Paul Bond, Tom Burnley, Renzhi Cao, Jianlin Cheng, Grzegorz Chojnowski, Kevin Cowtan, Ken A. Dill, Frank DiMaio, Daniel P. Farrell, James S. Fraser, Mark A. Herzik, Soon Wen Hoh, Jie Hou, Li-Wei Hung, Maxim Igaev, Agnel P. Joseph, Daisuke Kihara, Dilip Kumar, Sumit Mittal, Bohdan Monastyrskyy, Mateusz Olek, Colin M. Palmer, Ardan Patwardhan, Alberto Perez, Jonas Pfab, Grigore D. Pintilie, Jane S. Richardson, Peter B. Rosenthal, Daipayan Sarkar, Luisa U. Schäfer, Michael F. Schmid, Gunnar F. Schröder, Mrinal Shekhar, Dong Si, Abishek Singharoy, Genki Terashi, Thomas C. Terwilliger, Andrea Vaiana, Liguo Wang, Zhe Wang, Stephanie A. Wankowicz, Christopher J. Williams, Martyn Winn, Tianqi Wu, Xiaodi Yu, Kaiming Zhang, Helen M. Berman, and Wah Chiu. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nature Methods*, 18(2):156–164, 2021.

[29] Grigore Pintilie et al. Measurement of atom resolvability in cryo-EM maps with Q-scores, 2020.

[30] Tao Shen, Zhihang Hu, Siqi Sun, Di Liu, Felix Wong, Jiuming Wang, Jiayang Chen, Yixuan Wang, Liang Hong, Jin Xiao, Liangzhen Zheng, Tejas Krishnamoorthi, Irwin King, Sheng Wang, Peng Yin, James J. Collins, and Yu Li. Accurate rna 3d structure prediction using a language model-based deep learning approach. *Nature Methods*, 21(12):2287–2298, Dec 2024.

Appendix A

MEASURE OF FITNESS BETWEEN CRYO-EM MAP AND ATOMIC MODEL OF THE CURATED DATASET

Algorithm 1 Voxel-Level Labeling of RNA Cryo-EM Motif Segments

Require: Segmented motif density map M_{seg}, corresponding motif atoms S_{motif}
Ensure: Normalized density segment $\widetilde{M}_{\text{seg}}$ and labeled voxel map M_L
Load segmented cryo-EM density map M_{seg}
Extract voxel size, grid dimensions, and map origin from M_{seg}
Density normalization:
    Compute the 95th percentile of all non-zero density values in M_{seg}
    Scale density values by this percentile
    Clamp normalized density values to the range $[0, 1]$
Initialize an empty label grid M_L with the same dimensions as M_{seg}
Atomic-to-voxel mapping:
for each atom in S_{motif} **do**
    Retrieve atomic coordinates (x, y, z) in Å
    Convert (x, y, z) to voxel indices using the map origin and voxel size
    if the voxel index lies within the density grid **then**
        if atom is a backbone phosphorus atom **then**
            Assign the voxel to the backbone label
        else if atom is a ribose carbon or oxygen atom **then**
            Assign the voxel to the ribose sugar label
        else if atom is a nucleobase carbon, nitrogen, or oxygen atom **then**
            Assign the voxel to the nucleobase label
        end if
    end if
end for
Save $\widetilde{M}_{\text{seg}}$ and M_L with preserved voxel spacing and origin

Algorithm 2 Quality-Filtered RNA Motif Template Construction

Require: Motif MRC dataset $\mathcal{D}$, maximum resolution $R_{\max}$, minimum Q-score $Q_{\min}$, maximum samples per class $N_{\max}$

Ensure: Mean density template for each RNA motif class

for all motif map $m \in \mathcal{D}$ **do**
  Extract EMDB ID, PDB ID, and motif class from the filename.
  Retrieve the deposited map resolution from the resolution table or EMDB.
  if resolution $> R_{\max}$ **then**
    Discard m.
  end if
  Obtain the Q-score from the quality table or estimate it from density statistics.
  if Q-score $< Q_{\min}$ **then**
    Discard m.
  end if
  Read the density volume and reject empty or invalid maps.
  Assign the sample to its corresponding motif class.
end for
for all motif class c **do**
  Rank samples by Q-score.
  Retain at most $N_{\max}$ samples, preserving both quality and diversity.
  for all selected sample **do**
    Resample to the target voxel size.
    Center-crop to a fixed template size.
    Normalize the density using Z-score normalization.
  end for
  Compute the voxel-wise mean of all normalized samples.
  Save the resulting average density as the template for class c.
end for

Table A.1: Distribution of segmented data (total counts of atomic model and density map) in the dataset, by motif class and nucleotide category.

Class	**Category**	**Count**
Symmetric loops		
	1×1	6,078
	2×2	2,503
	3×3	2,425
	4×4	113
	5×5	244
	Subtotal	*11,363*
Asymmetric loops		
	1×2	3,614
	1×3	2,664
	1×4	1,687
	1×5	428
	2×3	2,472
	2×4	1,105
	2×5	580
	3×4	1,559
	3×5	49
	4×5	234
	Subtotal	*14,392*
Hairpins		
	hairpin3	4,174
	hairpin4	14,779
	hairpin5	23,449
	hairpin6	18,264
	hairpin7	3,768
	Subtotal	*64,434*
Bulges		
	bulge1	10,399
	bulge2	14,391
	bulge3	2,703
	bulge4	1,182
	bulge5	222
	Subtotal	*28,897*
Total		**119,086**

Table A.2: Distribution of CC_{mask} validation samples by motif class and map resolution

Motif – Resolution Å	**1.5–2.0**	**2.0–2.5**	**2.5–3.0**	**3.0–3.5**	**3.5–4.0**	**4.0–5.0**	**5.0–6.0**	**6.0–7.0**	**7.0–10.0**	**10.0–34.0**	**Total**
bulge1	0	0	0	6	0	36	25	62	264	378	771
bulge2	11	42	618	890	536	167	38	26	150	51	2529
bulge3	3	3	118	135	97	39	8	4	50	19	476
bulge4	1	3	41	79	45	12	2	4	6	0	193
bulge5	0	0	10	13	2	2	0	0	2	0	29
hairpin3	0	1	67	194	188	68	17	11	8	0	554
hairpin4	0	0	0	6	0	54	31	85	886	0	1062
hairpin5	0	26	384	1222	1173	421	111	59	68	12	3476
hairpin6	0	23	284	932	886	297	72	51	40	7	2592
hairpin7	0	0	0	4	0	16	18	30	95	105	268
1x2	13	96	252	407	226	93	30	16	60	38	1231
1x3	30	252	541	850	448	171	86	50	125	95	2648
1x4	15	153	339	572	313	128	47	33	50	27	1677
1x5	2	38	87	137	92	48	10	6	7	1	428
2x2	0	1	42	178	135	66	5	10	25	8	470
2x3	0	3	54	143	160	67	7	9	23	3	469
2x4	0	1	29	128	103	27	7	4	9	10	318
2x5	3	18	57	124	66	23	6	3	11	4	315
3x3	4	33	31	115	65	35	11	1	10	2	307
3x4	2	12	35	104	92	32	10	3	12	13	315
3x5	0	7	21	5	6	2	2	1	5	0	49
4x4	0	3	21	14	25	27	0	2	13	8	113
4x5	1	25	46	66	39	27	10	3	11	5	233
5x5	4	18	66	83	44	11	5	2	6	4	243
Total	89	758	3143	6407	4741	1869	558	475	1936	790	20766

Table A.3: Average cross correlation trends across motifs and resolution

Motif	**1.5–2.0**	**2.0–2.5**	**2.5–3.0**	**3.0–3.5**	**3.5–4.0**	**4.0–5.0**	**5.0–6.0**	**6.0–7.0**	**7.0–10.0**	**10.0–34.0**	**Overall**
1x2	0.846	0.806	0.818	0.742	0.702	0.593	0.589	0.562	0.508	0.402	0.721
1x3	0.809	0.785	0.793	0.749	0.700	0.617	0.556	0.611	0.492	0.413	0.717
1x4	0.860	0.851	0.824	0.783	0.746	0.658	0.634	0.618	0.466	0.399	0.761
1x5	0.911	0.851	0.820	0.796	0.775	0.686	0.631	0.630	0.640	0.488	0.780
2x2	–	0.885	0.820	0.704	0.678	0.611	0.604	0.688	0.529	0.503	0.680
2x3	–	0.630	0.844	0.739	0.715	0.672	0.639	0.616	0.428	0.288	0.711
2x4	–	0.137	0.806	0.747	0.699	0.652	0.659	0.658	0.404	0.488	0.706
2x5	0.849	0.832	0.810	0.785	0.758	0.600	0.613	0.542	0.687	0.600	0.762
3x3	0.857	0.823	0.723	0.758	0.672	0.601	0.591	0.451	0.495	0.571	0.710
3x4	0.878	0.843	0.809	0.763	0.729	0.597	0.597	0.551	0.425	0.623	0.724
3x5	–	0.819	0.797	0.815	0.748	0.376	0.355	0.704	0.575	–	0.736
4x4	–	0.714	0.746	0.609	0.640	0.612	–	0.677	0.482	0.375	0.615
4x5	0.803	0.734	0.745	0.727	0.644	0.597	0.656	0.494	0.594	0.517	0.686
5x5	0.851	0.851	0.821	0.797	0.736	0.610	0.470	0.557	0.395	0.369	0.763
bulge1	–	–	–	0.756	–	0.626	0.505	0.490	0.358	0.294	0.368
bulge2	0.786	0.706	0.765	0.689	0.608	0.514	0.434	0.527	0.255	0.292	0.641
bulge3	0.862	0.588	0.781	0.694	0.631	0.532	0.501	0.483	0.270	0.278	0.624
bulge4	0.447	0.641	0.663	0.626	0.620	0.551	0.451	0.513	0.405	–	0.616
bulge5	–	–	0.582	0.548	0.482	0.434	–	–	0.141	–	0.519
hairpin3	–	0.801	0.634	0.599	0.597	0.567	0.587	0.521	0.542	–	0.596
hairpin4	–	–	–	0.752	–	0.662	0.554	0.590	0.481	–	0.507
hairpin5	–	0.733	0.797	0.718	0.700	0.631	0.582	0.595	0.499	0.478	0.699
hairpin6	–	0.641	0.722	0.716	0.686	0.630	0.609	0.575	0.468	0.292	0.685
hairpin7	–	–	–	0.774	–	0.676	0.563	0.585	0.566	0.394	0.518
All motifs	**0.827**	**0.795**	**0.782**	**0.726**	**0.685**	**0.615**	**0.571**	**0.574**	**0.446**	**0.346**	**0.675**

Table A.4: Average Q-score correlation trend per motif and resolution

Motif – Resolution Å	**1.5-2.0**	**2.0-2.5**	**2.5-3.0**	**3.0-3.5**	**3.5-4.0**	**4.0-5.0**	**5.0-6.0**	**6.0-7.0**	**7.0-10.0**	**10.0-34.0**	**Overall**
bulge1		0.103	0.112	0.169	0.091	0.264	0.28	0.302	0.22	0.1	0.17
bulge2	0.794	0.692	0.699	0.568	0.439	0.371	0.258	0.339	0.108	0.069	0.514
bulge3	0.862	0.62	0.713	0.57	0.486	0.398	0.292	0.289	0.134	0.067	0.498
bulge4	0.467	0.6	0.569	0.478	0.419	0.317	0.243	0.313	0.105		0.455
bulge5			0.455	0.403	0.152	0.237			0.036		0.36
hairpin3		0.755	0.554	0.455	0.405	0.334	0.283	0.24	0.212		0.423
hairpin5		0.67	0.691	0.564	0.479	0.377	0.284	0.28	0.153	0.109	0.504
hairpin6		0.661	0.627	0.565	0.469	0.373	0.289	0.271	0.17	0.096	0.497
hairpin7				0.68		0.409	0.319	0.295	0.208	0.107	0.203
1x2	0.868	0.76	0.715	0.607	0.506	0.413	0.318	0.31	0.197	0.067	0.563
1x3	0.83	0.738	0.689	0.599	0.498	0.388	0.282	0.274	0.185	0.076	0.547
1x4	0.868	0.802	0.734	0.644	0.545	0.444	0.313	0.324	0.181	0.064	0.605
1x5	0.833	0.785	0.727	0.66	0.596	0.492	0.339	0.321	0.255	0.158	0.633
2x2	0.823	0.712	0.663	0.552	0.453	0.339	0.267	0.313	0.187	0.121	0.515
2x3	0.858	0.766	0.707	0.607	0.495	0.394	0.292	0.284	0.181	0.096	0.571
2x4	0.824	0.749	0.664	0.593	0.467	0.395	0.276	0.261	0.169	0.125	0.555
2x5	0.85	0.763	0.714	0.644	0.528	0.416	0.308	0.27	0.235	0.169	0.62
3x3	0.8	0.735	0.669	0.56	0.457	0.351	0.279	0.309	0.173	0.094	0.527
3x4	0.808	0.742	0.681	0.59	0.489	0.38	0.273	0.271	0.154	0.085	0.554
3x5		0.749	0.688	0.635	0.538	0.233	0.227	0.336	0.208		0.58
4x4		0.644	0.642	0.402	0.384	0.267		0.322	0.206	0.127	0.373
4x5	0.7	0.612	0.601	0.547	0.422	0.346	0.317	0.177	0.205	0.089	0.481
5x5	0.87	0.8	0.709	0.638	0.506	0.395	0.243	0.216	0.13	0.106	0.605
All motifs	0.831	0.744	0.673	0.573	0.474	0.377	0.289	0.293	0.178	0.089	0.524

Supplementary Algorithm S1. Training 3D CNN for RNA Motif Group Classification

Require: cryo-EM density maps, motif annotation CSV files
Ensure: Trained five-class motif classifier

```
Group 25 fine-grained motif labels into motif classes for each density map sample
Resample density map to target voxel size
Normalize density to zero mean and unit variance
Align label and density volumes by padding
Randomly select a labeled voxel as patch center
Extract 64 × 64 × 64 voxel patch at 0.5 Å/voxel (a 32 Å field of view; at inference, 32^3 cubes at 1.0 Å/voxel span the same extent)
Assign grouped motif class label
Initialize 3D CNN with three Conv3D-ReLU-MaxPool blocks
Initialize fully connected classifier layers
Set optimizer to Adam with learning rate 10^−3
Set loss function to cross-entropy
for each training epoch do
    for each mini-batch do
        Forward propagate density patches through network
        Compute classification loss
        Backpropagate gradients
        Update network parameters
    end for
    Evaluate validation accuracy
end for
Save trained model parameters
```

Table A.5: Blind-test confusion matrix for the 3D CNN model trained in Fold 3. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.826, Macro Specificity = 0.955.

True\Pred	sym	bul	hp	asymm	unk
sym	**72**	1	1	16	0
bul	1	**78**	11	0	0
hp	4	14	**62**	10	0
asymm	14	1	5	**69**	0
unk	0	0	0	0	**79**

Table A.6: Blind-test confusion matrix for the 3D CNN model trained in Fold 4. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.828, Macro Specificity = 0.956.

True\Pred	sym	bul	hp	asymm	unk
sym	**71**	0	5	14	0
bul	1	**77**	12	0	0
hp	3	13	**66**	8	0
asymm	16	0	4	**69**	0
unk	0	0	1	0	**78**

Table A.7: Blind-test confusion matrix for the 3D CNN model trained in Fold 5. Rows indicate ground-truth classes and columns indicate predicted classes. The blind test set contains 438 samples: 90 symmetric loops, 90 bulges, 90 hairpins, 89 asymmetric loops, and 79 unknown/background patches. Macro Sensitivity = 0.856, Macro Specificity = 0.963.

True\Pred	sym	bul	hp	asymm	unk
sym	**71**	1	2	16	0
bul	0	**85**	5	0	0
hp	2	15	**67**	6	0
asymm	9	0	5	**75**	0
unk	0	2	0	1	**76**

Table A.8: Cryo-EM maps and fitted atomic models from which the RNA motifs were extracted.

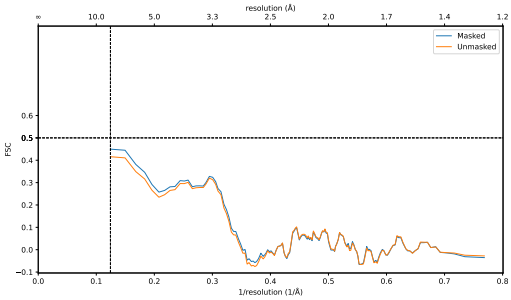
Regime	EMD ID	PDB	Biological complex	Res. (Å)
High	EMD-12217	7BL4	*In vitro* reconstituted 50S–ObgE–GMPPNP–RsfS particle	2.40
High	EMD-10076	6S0X	Erythromycin-resistant *Staphylococcus aureus* 70S ribosome (ΔR88 A89 uL22) in complex with erythromycin	2.425
Mid	EMD-4477	6Q98	tmRNA–SmpB bound in the P site of the *Escherichia coli* 70S ribosome	4.30
Mid	EMD-11768	7AFK	Bacterial 30S ribosomal subunit assembly complex, state D (head domain)	4.90
Low	EMD-2810	4D5L	Mammalian 80S ribosome termination complex with Cricket paralysis virus (CrPV) IRES RNA in the translocated state and eRF1	8.7

Appendix B

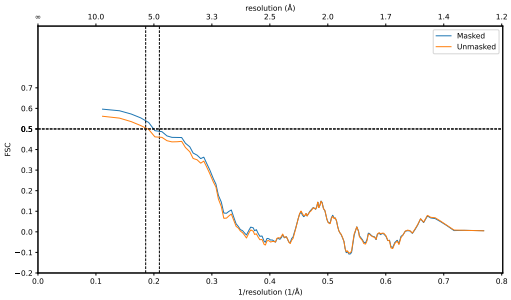
MAP–TO–MODEL FSC ANALYSIS

Table B.1: Map–to–model FSC curves indexed by map resolution regime and randomly selected RNA motif type. Each cell gives the random motif instance under the motif group and a reference to its FSC plot.

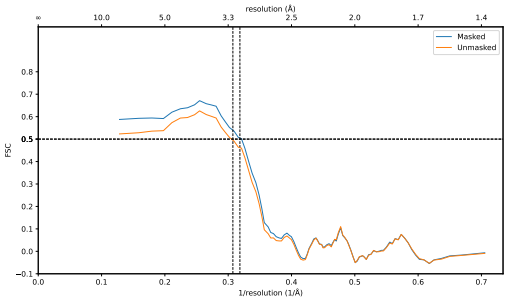
Res. \ Motif	Symmetric	Asymmetric	Hairpin	Bulge
High	3×3 (Fig. B.1a)	3×4 (Fig. B.1b)	hairpin6 (Fig. B.1c)	bulge4 (Fig. B.1d)
Mid	2×2 (Fig. B.2a)	1×3 (Fig. B.2b)	hairpin6 (Fig. B.2c)	bulge3 (Fig. B.2d)
Low	3×3 (Fig. B.3a)	1×5 (Fig. B.3b)	hairpin5 (Fig. B.3c)	bulge4 (Fig. B.3d)



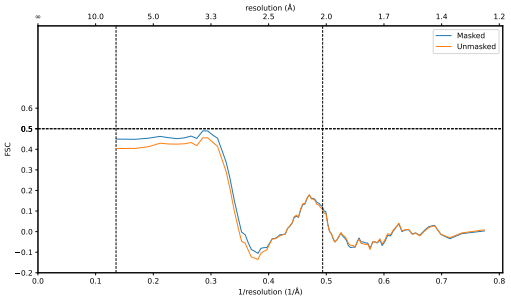
(a) Symmetric internal loop (3×3).



(b) Asymmetric internal loop (3×4).

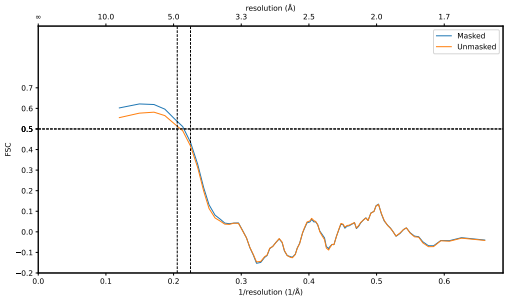


(c) Hairpin loop.

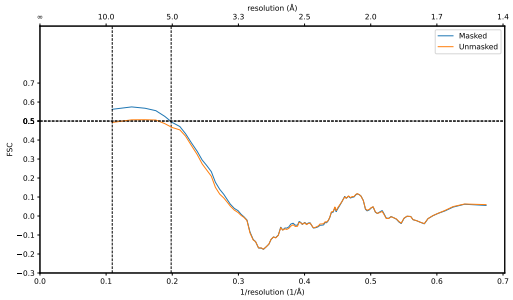


(d) Bulge.

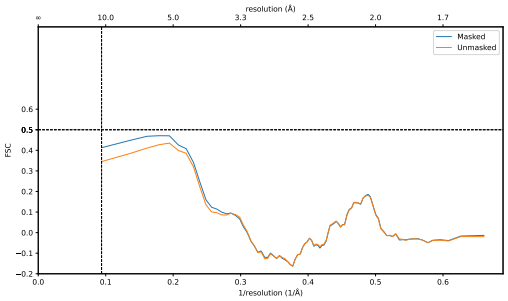
Figure B.1: **High-resolution regime.** Map–to–model FSC for RNA motifs extracted from EMD-12217 / PDB 7BL4 (*in vitro* reconstituted 50S–ObgE–GMPPNP–RsfS particle, 2.40 Å); the hairpin (c) is from EMD-10076 / PDB 6S0X (erythromycin-resistant *Staphylococcus aureus* 70S ribosome, 2.425 Å).



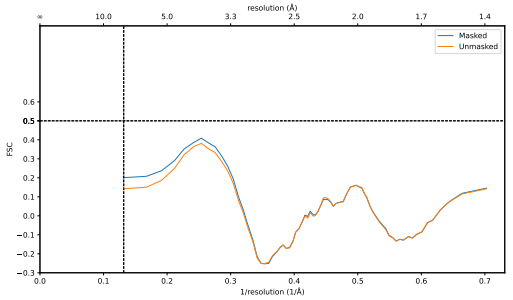
(a) Symmetric internal loop (2×2).



(b) Asymmetric internal loop (1×3).

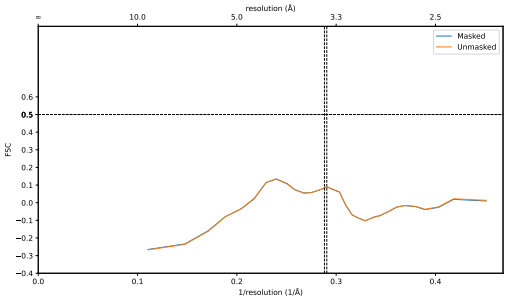


(c) Hairpin loop.

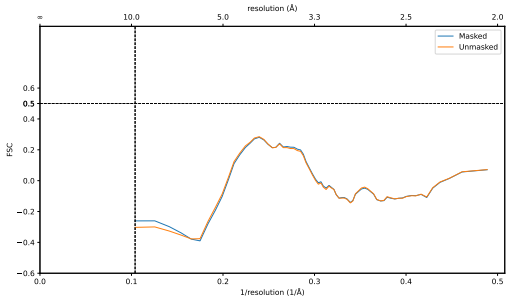


(d) Bulge.

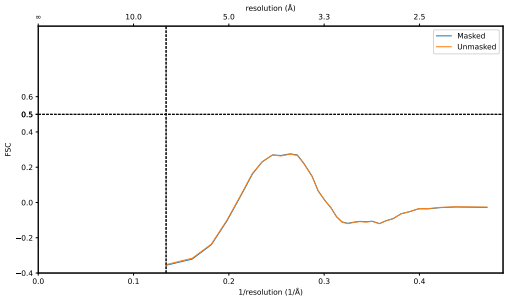
Figure B.2: **Mid-resolution regime.** Map–to–model FSC for RNA motifs from EMD-4477 / PDB 6Q98 (tmRNA–SmpB bound in the P site of the *Escherichia coli* 70S ribosome, 4.30 Å); the bulge (d) is from EMD-11768 / PDB 7AFK (bacterial 30S ribosomal subunit assembly complex, state D, head domain, 4.90 Å).



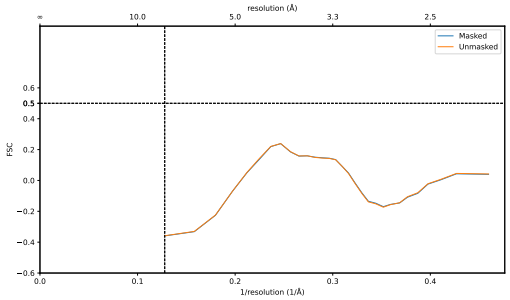
(a) Symmetric internal loop (3×3).



(b) Asymmetric internal loop (1×5).



(c) Hairpin loop.



(d) Bulge.

Figure B.3: **Low-resolution regime.** Map–to–model FSC for RNA motifs extracted from EMD-2810 / PDB 4D5L (mammalian 80S ribosome termination complex with Cricket paralysis virus (CrPV) IRES RNA in the translocated state and eRF1, 8.7 Å).