



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR MATHEMATIK, INFORMATIK
UND NATURWISSENSCHAFTEN

DISSERTATION

Computational characterization of macromolecular interfaces in AI-predicted complex structures

Luca Ragna Genz

Integrative Virology

Department of Chemistry

Faculty of Mathematics, Informatics and Natural Sciences

Universität Hamburg

Hamburg, Germany

A thesis submitted for the degree of

Doctor rerum naturalium (Dr. rer. nat.)

Computational characterization of macromolecular interfaces in AI-predicted complex structures

Dissertation submitted by: Luca Ragna Genz

Date of Submission: 26.05.2026

Date of Disputation: 28.08.2026

Supervisor: Prof. Dr. Maya Topf, Universität Hamburg

Co-supervisor: Prof. Dr. Kay Grünewald, Universität Hamburg

Committee:

1st Prof. Dr. Daniel Wilson, Universität Hamburg

2nd Prof. Dr. Maya Topf, Universität Hamburg

3rd Prof. Dr. Kay Grünewald, Universität Hamburg

4th Prof. Dr. Pietro Scaturro, Universität Hamburg

5th Dr. Benjamin Vollmer, Universität Hamburg

Reviewers:

1st Prof. Dr. Maya Topf, Universität Hamburg

2nd Prof. Dr. Daniel Wilson, Universität Hamburg

3rd Dr. Damien Devos, Center for Infection & Immunity of Lille (CIIL), Institut Pasteur de Lille

Universität Hamburg, Hamburg, Germany

Faculty of Mathematics, Informatics and Natural Sciences

Department of Chemistry

Integrative Virology

The work described in this thesis was conducted between November 2022 and May 2026. The work was performed at the Centre for Structural Systems Biology (CSSB) in Hamburg and funded by the Leibniz Institute for Virology (LIV).

Abstract

Protein-protein, protein-nucleic acid, and nucleic acid-only complexes mediate essential cellular processes and represent key targets for structure-guided drug discovery. Recent advances in AI-based structure prediction, including AlphaFold2, AlphaFold3, and ColabFold, have substantially expanded the availability of structural models while simultaneously creating a need for computational frameworks capable of reliably characterizing and assessing predicted complex interfaces across diverse biomolecular interaction types. This thesis presents three computational approaches that address this challenge.

First, PICKLUSTER is introduced as a tool for the spatial subdivision of protein-protein interfaces into geometrically distinct sub-interfaces, implemented as both a UCSF ChimeraX plug-in and a standalone command-line tool. Based on distance-dependent clustering of heavy-atom contacts, the method enables independent analysis of spatially separated interface regions. It integrates a range of interface-specific confidence metrics, including ipLDDT and iPAE, to support sub-interface-level quality assessment of AlphaFold- and ColabFold-based predictions. Application of PICKLUSTER to a predicted *Epstein-Barr virus* protein complex demonstrates how interfaces can be decomposed into sub-interfaces that differ substantially in prediction confidence, thereby enabling targeted prioritization for downstream analyses. Large-scale analysis further reveals characteristic patterns in the spatial organization of protein-protein interfaces and identifies extreme cases, such as a complex from *Porphyridium purpureum* containing 4,844 sub-interfaces.

Second, a systematic benchmarking study is conducted to evaluate commonly used confidence metrics for the assessment of heterodimeric protein complex predictions generated by ColabFold with templates, ColabFold without templates, and AlphaFold3, using DockQ as a reference measure on a curated benchmark dataset. The analysis identifies the most reliable metric for reliable interface assessment, highlights prediction method-dependent differences in performance, and establishes binary classification thresholds for each confidence score, thereby providing a practical framework for the interpretation and evaluation of predicted protein complexes. Based on these findings, C2Qscore is introduced as a combined metric that integrates complementary features to improve robustness in model quality evaluation. C2Qscore is validated on independent X-ray crystallography and cryo-EM-derived test sets and integrated into PICKLUSTER v2.0 as an interactive module.

Third, ProNA3D extends the PICKLUSTER framework to protein-nucleic acid and nucleic acid-only complexes, providing a ChimeraX plug-in and a command-line tool that integrates sub-interface detection, secondary-structure topology visualization, interface contact networks, and AlphaFold3 confidence metrics for predictions of nucleic acid-containing complexes. The utility of ProNA3D is demonstrated on a diverse range of nucleic acid-containing complexes solved by X-ray crystallography or cryo-EM, as well as computationally predicted models. For example, the tool identifies interaction hubs

in the interface of a dimeric hammerhead ribozyme and a trimeric HIV-1 RNA-human antibody complex. In addition, the method supports the investigation of flexible regions in cryo-EM density maps, thereby facilitating localized model validation. Large-scale application to nucleic acid-containing complexes further reveals distinct structural and connectivity patterns across interaction types, including differences in flexibility and interaction organization between nucleic acid-only and protein-nucleic acid interfaces. This analysis additionally enables the identification of specialized interaction modes, such as those observed in the complex formed by the CRISPR-associated protein Csm4 and crRNA. Finally, analysis of a curated subset of viral structures identifies characteristic features of viral nucleic acid-containing complexes.

The results presented in this thesis highlight that macromolecular interfaces are spatially heterogeneous, making sub-interface-level analysis essential for reliable structural interpretation. Taken together, the developed methods and analyses establish an integrated computational framework for a confidence-aware interface characterization across a broad spectrum of biomolecular interactions, bridging the gap between large-scale AI-based structure prediction and reliable structural interpretation.

Zusammenfassung

Protein-Protein-, Protein-Nukleinsäure- und Nukleinsäure-Nukleinsäure-Komplexe vermitteln essentielle zelluläre Prozesse und stellen wichtige Zielstrukturen für die strukturbasierte Wirkstoffentwicklung dar. Jüngste Fortschritte in der KI-basierten Strukturvorhersage, darunter AlphaFold2, AlphaFold3 und ColabFold, haben die Verfügbarkeit von Strukturmodellen erheblich erweitert und zugleich einen erhöhten Bedarf an Methoden geschaffen, die eine zuverlässige Charakterisierung vorhergesagter Schnittstellen in verschiedenen biomolekularen Interaktionstypen ermöglichen. Diese Dissertation stellt drei Ansätze vor, die diesen Herausforderungen begegnen.

Zuerst wird PICKLUSTER als Softwareprogramm zur räumlichen Unterteilung von Protein-Protein-Schnittstellen in geometrisch separierte Teilschnittstellen vorgestellt, das sowohl als UCSF ChimeraX-Plugin als auch als eigenständiges Befehlszeilenprogramm implementiert ist. Basierend auf einer abstandsabhängigen Gruppierung von Kontakten zwischen Schweratomen ermöglicht PICKLUSTER die Analyse räumlich getrennter Schnittstellenbereiche. Es integriert eine Reihe von schnittstellenspezifischen Qualitätsmetriken, darunter ipLDDT und iPAE, um die Bewertung von AlphaFold-basierten Vorhersagen auf Teilschnittstellenebene zu unterstützen. Die Anwendung von PICKLUSTER auf einen vorhergesagten Proteinkomplex des *Epstein-Barr-Virus* zeigt, dass einzelne Schnittstellen mehrere Regionen enthalten können, die sich hinsichtlich ihrer Vorhersagequalität erheblich voneinander unterscheiden. Dies ermöglicht die Priorisierung von Bereichen innerhalb der Interaktionsschnittstelle für die strukturbasierte Wirkstoffentwicklung. Eine groß angelegte Analyse identifiziert darüber hinaus charakteristische Merkmale in der räumlichen Organisation von Protein-Protein-Schnittstellen und zeigt Extremfälle auf, wie beispielsweise in einem Komplex aus *Porphyridium purpureum*, der 4.844 Teilschnittstellen enthält.

Zweitens wird eine systematische Benchmark-Studie durchgeführt, in der gängige Konfidenzmetriken zur Qualitätsbewertung von Vorhersagen heterodimerer Proteinkomplexe basierend auf einem selektierten Proteinstrukturdatensatz bewertet werden, wobei die Metrik DockQ als Referenz dient. Zur Vorhersage der Strukturen werden drei verschiedene Methoden (ColabFold mit Zugriff auf die PDB, ColabFold ohne Zugriff auf die PDB und AlphaFold3) genutzt. Die Analyse identifiziert Konfidenzmetriken für eine zuverlässige Bewertung vorhergesagter Interaktionsflächen, beleuchtet Qualitätsunterschiede zwischen den Vorhersagemethoden und bestimmt binäre Klassifikationsschwellenwerte für jede Metrik, wodurch eine Anleitung für die Bewertung vorhergesagter Proteinkomplexe bereitgestellt wird. Auf dieser Grundlage wird C2Qscore als kombinierte Metrik eingeführt, die komplementäre Merkmale integriert und somit eine robustere Modellqualitätsbewertung ermöglicht. C2Qscore wird an unabhängigen Röntgenkristallografie- und Kryo-EM-basierten Strukturdatensätzen validiert und als interaktives Modul in PICKLUSTER v2.0 integriert.

Drittens weitet ProNA3D den PICKLUSTER-Ansatz auf Protein-Nukleinsäure-Komplexe und reine Nukleinsäurekomplexe aus und stellt ein ChimeraX-Plugin sowie ein Befehlszeilentool bereit, die die Erkennung von Teilschnittstellen, die Visualisierung der Sekundärstruktur-Topologie, Schnittstellen-Kontaktnetzwerke und AlphaFold3-Konfidenzmetriken für Vorhersagen zu nukleinsäurehaltigen Komplexen integrieren. Die Nützlichkeit von ProNA3D wird anhand spezifischer nukleinsäurehaltiger Komplexe, die mittels Röntgenkristallographie oder Kryo-EM aufgeklärt wurden, sowie anhand vorhergesagter Modelle demonstriert. Beispielsweise identifiziert ProNA3D Interaktionsknotenpunkte in einem dimeren Hammerhead-Ribozym sowie in einem trimeren Komplex aus HIV-1-RNA und einem humanen Antikörper. Darüber hinaus unterstützt die Methode die Untersuchung flexibler Regionen in Kryo-EM-Dichtekarten und vereinfacht so die lokalisierte Modellvalidierung. Die groß angelegte Anwendung auf Strukturen von nukleinsäurehaltigen Komplexen deckt zudem unterschiedliche Struktur- und Konnektivitätsmuster über Interaktionstypen hinweg auf, einschließlich Unterschiede in der Flexibilität und Interaktionsorganisation zwischen reinen Nukleinsäure- und Protein-Nukleinsäure-Schnittstellen. Diese Analyse ermöglicht zudem die Identifizierung spezialisierter Interaktionsmodi, wie beispielsweise in einem Komplex, der aus dem CRISPR-assoziierten Protein Csm4 und crRNA besteht. Darüber hinaus identifiziert die Analyse eines spezialisierten Datensets charakteristische Merkmale viraler Protein-Nukleinsäure-Komplexe.

Die in dieser Arbeit vorgestellten Ergebnisse zeigen, dass makromolekulare Schnittstellen räumlich heterogen sind, weshalb die Analyse von Teilschnittstellen für eine zuverlässige Strukturinterpretation unerlässlich ist. Insgesamt etablieren die in dieser Arbeit entwickelten Methoden und Analysen einen integrierten rechnergestützten Ansatz zur konfidenzbasierten Charakterisierung biomolekularer Interaktionsschnittstellen über ein breites Spektrum biomolekularer Komplexe hinweg und tragen damit dazu bei, die Lücke zwischen KI-basierter Strukturvorhersage und zuverlässiger Strukturinterpretation zu schließen.

Contents

List of Publications	iv
List of Figures	v
List of Tables	viii
List of Abbreviations	xi
1 Introduction	1
1.1 General background and motivation	2
1.1.1 Macromolecular interactions as the foundation of biology . . .	3
1.1.2 The structural basis of biomolecular function	4
1.1.3 Protein-protein and nucleic acid-containing complexes in disease and drug discovery	5
1.2 Protein complex structure determination	5
1.2.1 The structure of proteins	6
1.2.2 Structural principles of protein-protein interfaces	7
1.2.3 Experimental structure determination of proteins and protein complexes	10
1.2.4 Deposition and access of structural data	12
1.2.5 Limitations of X-ray crystallography and cryo-EM	13
1.3 Protein complex structure prediction	14
1.3.1 Computational protein structure prediction	14
1.3.2 Computational prediction of multi-chain complexes	17
1.4 Analysis and evaluation of protein-protein interfaces	18
1.4.1 Molecular visualization of protein structures and complexes . .	19
1.4.2 Tools for protein complex interface identification, analysis, and clustering	19
1.4.3 Reference-based quality metrics	20
1.4.4 Scoring metrics for AI-generated protein complex predictions .	21
1.4.5 Benchmarking initiatives in structural model evaluation	23
1.4.6 Motivation for PICKLUSTER	23
1.4.7 Motivation for the benchmarking of scoring metrics and C2Qscore	24
1.5 Beyond proteins: Expanding to nucleic acid-containing complexes . . .	25
1.5.1 Biological roles and structural features of nucleic acid-containing complexes	25
1.5.2 Challenges in structural determination and prediction of nucleic acid-containing complexes	28

1.5.3	Computational tools for the analysis of nucleic acid-containing complexes	29
1.5.4	Motivation for ProNA3D	30
2	Aim	32
3	Results	34
3.1	PICKLUSTER: A Protein-Interface Clustering and Analysis Plug-in for UCSF ChimeraX	35
3.2	Benchmarking PICKLUSTER on Large-Scale Protein Complex Data . .	37
3.3	Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions	39
3.3.1	Benchmarking the prediction accuracy of ColabFold with templates, ColabFold without templates, and AlphaFold3	39
3.3.2	Influence of the heavy atom distance parameter for interface calculation	43
3.3.3	Distributions and correlation of individual prediction-based scores	43
3.3.4	Correlation of the individual prediction-based scores with DockQ	44
3.3.5	Determinants of high- and low-quality models	46
3.3.6	Combined scoring and testing	49
3.3.7	Evaluating the scores on dimers from larger assemblies	50
3.4	ProNA3D: Distance-Based Analysis of Nucleic Acid and Protein Interfaces	52
3.4.1	X-ray structure of a protein-RNA complex: <i>Acanthamoeba polyphaga</i> mimivirus RNA guided nuclease-RNA	54
3.4.2	X-ray structure of an RNA-only complex: hammerhead ribozyme	56
3.4.3	Predicted protein-RNA complex: human Fab antibody-HIV-1 RNA	58
3.4.4	Predicted protein-DNA complex: Cre recombinase from <i>Punavirus P1</i> bound to its target DNA	60
3.4.5	Cryo-EM application of ProNA3D	62
3.5	Systematic analysis of protein-nucleic acid and nucleic acid-only interfaces using ProNA3D	64
3.5.1	Interface size distributions across nucleic acid-containing complexes	64
3.5.2	Normalized B-factor analysis of interface regions	68
3.5.3	Maximum connectivity distributions for different interface classes	69
3.5.4	Individual residue- and nucleotide-level connectivity for different interface classes	74
3.5.5	Occurrence distributions of residue- and nucleotide-types for different interface classes	76
3.5.6	Nucleic acid-containing interfaces in viral complexes	80
4	Discussion	85
4.1	PICKLUSTER: spatial decomposition of protein-protein interfaces . . .	86
4.2	Large-scale analysis of protein-protein interface properties	87
4.3	Quality assessment of predicted protein complex models	88
4.3.1	Performance hierarchy of individual scoring metrics	89
4.3.2	Assessment of prediction quality across AlphaFold3 and ColabFold	90
4.3.3	C2Qscore	91

4.4	ProNA3D: extending interface analysis to protein-nucleic acid and nucleic acid-only complexes	92
4.4.1	Detection of interaction hotspots	93
4.4.2	Confidence-guided interface analysis of AlphaFold3-predicted protein-nucleic acid complexes	94
4.4.3	Interface-based density zoning for local quality assessment in cryo-EM structures	94
4.5	Systematic analysis of nucleic acid-containing interfaces in the PDB . .	95
4.5.1	Interface size distributions across nucleic acid-containing complex classes	96
4.5.2	Protein binding stabilizes nucleic acid interface flexibility . . .	96
4.5.3	Connectivity patterns in nucleic-acid containing complexes . .	97
4.5.4	Structural features of viral nucleic acid-containing interfaces .	98
4.6	Limitations	99
4.6.1	Limitations of the PICKLUSTER and ProNA3D algorithm . . .	99
4.6.2	C2Qscore	100
4.6.3	Limitations of the large-scale datasets and analyses	101
4.6.4	General considerations	102
5	Conclusions and Outlook	103
5.1	Quality assessment of predicted protein-nucleic acid and nucleic acid-only complexes	105
5.2	Nucleic acid-containing and protein-protein complexes as targets for structure-guided drug design	106
5.3	Interface property profiles for molecular class annotation	106
5.4	Structural determinants of prediction quality in nucleic acid-containing complexes	107
6	Methods	108
6.1	Tool development: PICKLUSTER	109
6.1.1	Design and implementation	109
6.1.2	Interface calculation and clustering algorithm	110
6.1.3	Additional features and scoring metrics	111
6.1.4	Dataset for the large-scale analysis of protein-protein interfaces	112
6.2	Assessing scoring metrics for protein complex predicitions	112
6.2.1	Benchmark set and X-ray crystallography test set construction (Figure 6.1)	113
6.2.2	Model generation	114
6.2.3	Reference-based score: DockQ	115
6.2.4	Prediction-based scores	115
6.2.5	Combined scoring: C2Qscore	115
6.2.6	Pearson correlation coefficient	116
6.2.7	ROC curves and AUC	116
6.2.8	Accuracy	116
6.2.9	Precision	116
6.2.10	Wilcoxon signed-rank test	116
6.3	ProNA3D	117
6.3.1	General functions and implementation	117

6.3.2	Interface Extraction	117
6.3.3	Topology analysis	118
6.3.4	Two-dimensional interaction networks	118
6.3.5	AlphaFold3 scoring	119
6.3.6	Cryo-EM map analysis	119
6.3.7	Dataset curation for large-scale PDB analysis	119
6.3.8	Viral subset	120
6.3.9	B-factor normalization	120
6.3.10	Connectivity analysis	120
6.3.11	Mann-Whitney U test	121
6.4	Data and software availability	121
A	Additional Material	122
A.1	Supplemental figures	122
A.2	Supplemental tables	130
A.3	Supplemental text	143
B	Software and Tools	144
C	Hazardous Substances	146
	References	147

List of Publications

The following publications form the basis of this dissertation. Material adapted from these works is used in chapter 3, chapter 4, chapter 6, and Appendix A.

Research articles

- L. R. Genz, T. Mulvaney, S. Nair, and M. Topf (2023). PICKLUSTER: a protein-interface clustering and analysis plug-in for UCSF ChimeraX. In: *Bioinformatics* 39.11, btad629. DOI: 10.1093/bioinformatics/btad629.
- L. R. Genz, S. Nair, N. Nagar, and M. Topf (2025a). Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions. In: *Protein Sci.* 34.11, e70327. DOI: 10.1002/pro.70327.
- L. R. Genz and M. Topf (2026). ProNA3D: Distance-based analysis of nucleic acid-containing interfaces. In: *bioRxiv*. DOI: 10.64898/2026.04.16.719043.

Review articles

- L. R. Genz, S. Nair, A. Sweeney, and M. Topf (2025b). Drug targeting of protein-nucleic acid interactions. In: *Curr. Opin. Struct. Biol.* 95, 103165. DOI: 10.1016/j.sbi.2025.103165.

List of Figures

1.1	A collection of macromolecular complexes formed by protein-protein interactions.	3
1.2	The four hierarchical levels of protein structure.	6
1.3	Schematic representation of the major non-covalent interactions in protein structures.	8
1.4	The organization of hotspots within protein-protein interfaces.	9
1.5	Comparison of X-ray crystallography, cryo-EM, and NMR spectroscopy, highlighting their experimental principles, methodological advantages, and limitations.	12
1.6	Growth and composition of the Protein Data Bank.	13
1.7	Schematic overview of the homology modeling workflow.	15
1.8	Overview of the AlphaFold2 architecture and prediction confidence metrics.	16
1.9	Backbone agreement with experiment across 14 CASP rounds.	18
1.10	Overview of nucleic acid building blocks, nucleotide composition, and canonical Watson-Crick base pairs.	27
1.11	Overview of computational tools for analyzing the structure of nucleic acid-containing complexes.	30
3.1	PICKLUSTER applied to the dimeric complex of BDRF1 and BFRF1A . .	36
3.2	Analysis of a large-scale protein complex dataset using PICKLUSTER. .	38
3.3	Distribution of quality scores, overlap of DockQ classifications, and detailed target analysis for models predicted by CF-T, CF-F, and AF3. .	42
3.4	Distributions and correlations of model quality scores for the CF-T predictions.	44
3.5	Pearson correlation coefficients of individual scores with DockQ and representative structural predictions.	46
3.6	Optimal classification cutoffs, ROC curves, and representative gray zone predictions.	48
3.7	Unweighted combined score and C2Qscore evaluated on the X-ray crystallography test set.	50
3.8	Performance of the unweighted combined score and C2Qscore on dimers derived from cryo-EM assemblies of higher stoichiometry.	52
3.9	Overview of ProNA3D for integrated interface analysis within UCSF ChimeraX.	54
3.10	Interface modes provided by ProNA3D	55
3.11	Effect of the heavy-atom distance cutoff (d_{HA}) and topological analysis of a dimeric hammerhead ribozyme RNA complex (PDB ID: 2OEU) using ProNA3D.	57

3.12	ProNA3D-based analysis of the AlphaFold3 model for the CASP16 target M1296 in comparison with its experimentally resolved structure (PDB ID: 9C75).	59
3.13	ProNA3D-based analysis of an AlphaFold3 prediction of the Cre recombinase from <i>Punavirus P1</i> in complex with its target DNA.	61
3.14	Cryo-EM-focused ProNA3D analysis of the ATP-dependent DNA/RNA helicase DHX36 in complex with G-quadruplex RNA (PDB ID: 8VV2, EMD-43547)	63
3.15	Interface size distributions across nucleic acid-containing complexes.	67
3.16	Normalized mean B-factor distributions of interface regions across interaction types in X-ray crystallographic structures.	69
3.17	Distribution of maximum connectivity across nucleic acid-containing interface types.	71
3.18	Examples illustrating low- and high-connectivity interface outliers.	73
3.19	Per-nucleotide type connectivity (A: blue; U/T: lightblue; G: orange; C: lightorange) across nucleic acid-containing interface types.	75
3.20	Per-residue type connectivity across nucleic acid-containing interface types.	76
3.21	Per-nucleotide type occurrence (A: blue; U/T: lightblue; G: orange; C: lightorange) across nucleic acid-containing interface types.	78
3.22	Per-residue type occurrence across nucleic acid-containing interface types.	79
3.23	Interface size distributions across nucleic acid-containing complexes in the viral subset.	81
3.24	Normalized mean B-factor distributions of interface regions across interaction types in X-ray crystallographic structures in the viral dataset.	82
3.25	Distribution of maximum connectivity across nucleic acid-containing interface types in the viral subset.	84
6.1	Curation pipeline for the benchmark and the X-ray test set.	114
A.1	BALF3 (purple), BDRF1 (salmon), and BFRF1A (green) from <i>Epstein-Barr Virus</i> modelled as a trimeric complex with ColabFold.	122
A.2	Venn-diagrams of DockQ classifications for CF-T, CF-F and AF3.	123
A.3	Scatter plots of ipLDDT and iPAE versus DockQ across distance thresholds ranging from 5 to 9 Å.	124
A.4	Distributions and correlations of model quality scores for the CF-F predictions.	125
A.5	Distributions and correlations of model quality scores for the AF3 predictions.	126
A.6	ipTM and <i>model confidence</i> (mc) versus DockQ for the benchmark dataset across CF-T, CF-F, and AF3 predictions.	126
A.7	Scatter plots of ipLDDT, iPAE, pTM, ipTM, <i>model confidence</i> (mc), pDockQ2, and VoroIF against DockQ for the test set across CF-T, CF-F, and AF3 predictions.	127
A.8	Scatter plots of the unweighted combined score and C2Qscore for the large assembly test set across CF-T, CF-F, and AF3 predictions.	128

A.9	Distribution of DockQ and C2Qscore values for CF-T (blue), CF-F (green), and AF3 (orange) predictions corresponding to cases with two or three potential configurations.	129
A.10	Conserved nucleotide flipping associated with high-connectivity interactions in CRISPR-associated protein Csm4 complexes.	130

List of Tables

1.1	DockQ classification for protein complex models based on CAPRI criteria (Basu and Wallner 2016).	21
A.1	Prediction-based scores used for the assessment of the structural models. Adapted from Genz et al. (2025a).	130
A.2	Accuracy and precision for individual scores for CF-T, CF-F and AF3. Binary classification: incorrect vs. (high, medium, acceptable). The used thresholds are reported in Figure 3.6. The calculations are based on a binary classification: incorrect vs. (high, medium, acceptable).	131
A.3	Accuracy and precision for the gray zone of ipTM ($0.6 \leq \text{ipTM} \leq 0.8$) for CF-T, CF-F and AF3. The used threshold is reported in Figure 3.6A. The calculations are based on a binary classification: incorrect vs. (high, medium, acceptable).	132
A.4	Pearson correlation of all approaches used to determine a combined score for CF-T, CF-F and AF3.	132
A.5	Weights for C2Qscore for CF-T, CF-F and AF3.	132
A.6	Bias for the unweighted combined score and C2Qscore for CF-T, CF-F and AF3.	132
A.7	Maximal values for the unweighted combined score and C2Qscore for CF-T, CF-F and AF3.	132
A.8	P-values for the Wilcoxon signed-rank test of C2Qscore against the unweighted combination of all scores.	133
A.9	Determined thresholds for the unweighted combined score and C2Qscore for distinguishing models of low (incorrect DockQ) and high (high, medium, acceptable DockQ) quality for CF-T, CF-F and AF3.	133
A.10	Interface size statistics, sorted by interface type. For protein-nucleic acid interfaces, protein and nucleic acid sides are listed separately. n denotes the number of sub-interfaces; σ denotes the standard deviation.	133
A.11	Normalized B-factor statistics, restricted to X-ray crystallographic structures, stratified by interface type. B-factors are normalized per structure using min-max scaling. n denotes the number of sub-interfaces; σ denotes the standard deviation.	133
A.12	Maximum connectivity across all interface types. Shown are the number of sub-interfaces analyzed (n) and the distribution of maximal connectivity values, reported as median, mean, and standard deviation σ	134
A.13	Mann-Whitney U (Wilcoxon rank-sum) test for maximum connectivity interface type populations. P-values are Benjamini-Hochberg corrected for multiple testing ($n = 10$ pairwise comparisons).	134

A.14	Connectivity statistics by nucleotide type for DNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation. .	134
A.15	Connectivity statistics by nucleotide type for RNA-RNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation. .	134
A.16	Connectivity statistics by nucleotide type (RNA) for RNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.	135
A.17	Connectivity statistics by nucleotide type (DNA) for RNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.	135
A.18	Connectivity statistics by nucleotide type (DNA) for protein-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.	135
A.19	Connectivity statistics by residue type (protein) for protein-DNA interfaces. n denotes the number of residues; σ denotes the standard deviation.	136
A.20	Connectivity statistics by nucleotide type (RNA) for protein-RNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.	136
A.21	Connectivity statistics by residue type (protein) for protein-RNA interfaces. n denotes the number of residues; σ denotes the standard deviation.	137
A.22	Interface size statistics for the viral subset, stratified by interface type. For protein-nucleic acid interfaces, protein and nucleic acid sides are listed separately. n denotes the number of sub-interfaces; σ denotes the standard deviation.	138
A.23	Normalized B-factor statistics for the viral subset, restricted to X-ray crystallographic structures, stratified by interface type. B-factors are normalized per structure using min-max scaling. n denotes the number of sub-interfaces included in the B-factor analysis; σ denotes the standard deviation.	138
A.24	Maximum connectivity statistics for the viral subset, stratified by interface type. Maximum connectivity is defined as the highest number of inter-partner contacts formed by any single residue or nucleotide at a given sub-interface. n denotes the number of sub-interfaces; σ denotes the standard deviation.	139
A.25	Per-nucleotide mean connectivity at DNA-DNA interfaces in the viral subset. Connectivity is defined as the number of contacts formed by each nucleotide with its interface partner. n denotes the number of nucleotides; σ denotes the standard deviation.	139
A.26	Per-nucleotide mean connectivity at RNA-RNA interfaces in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.	139
A.27	Per-nucleotide mean connectivity at RNA-DNA interfaces (RNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.	139

A.28	Per-nucleotide mean connectivity at RNA-DNA interfaces (DNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.	140
A.29	Per-nucleotide mean connectivity at protein-DNA interfaces (DNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.	140
A.30	Per-nucleotide mean connectivity at protein-RNA interfaces (RNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.	140
A.31	Per-residue mean connectivity at protein-DNA interfaces (protein side) in the viral subset. n denotes the number of residues; σ denotes the standard deviation.	141
A.32	Per-residue mean connectivity at protein-RNA interfaces (protein side) in the viral subset. n denotes the number of residues; σ denotes the standard deviation.	142

List of Abbreviations

AF3	AlphaFold3
AI	artificial intelligence
ATP	adenosine triphosphate
AU	asymmetric unit
AUC	area under the curve
BA	biological assembly
BSA	buried surface area
C2Qscore	Combined Complex Quality score
CAMEO	Continuous Automated Model Evaluation
CAPRI	Critical Assessment of Predicted Interactions
CASP	Critical Assessment of Structure Prediction
CF-F	ColabFold without templates
CF-T	ColabFold with templates
COCOMAPS	bioCOMplexes COnTact MAPS
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
cryo-EM	cryogenic electron microscopy
CSV	comma-separated values
DNA	deoxyribonucleic acid
DockQ	protein docking quality score
dsRNA	double-stranded RNA
DSSP	Dictionary of Secondary Structure of Proteins
DSSR	Defining the Secondary Structures of RNA
EBV	Epstein-Barr virus
EMA	Estimation of Model Accuracy
EMDB	Electron Microscopy Data Bank
FDA	Food and Drug Administration
F_{nat}	fraction of native contacts
$F_{\text{non-nat}}$	fraction of non-native contacts
FN	false negative

FSC	Fourier Shell Correlation
FP	false positive
FPR	false positive rate
GDT_TS	Global Distance Test Total Score
GNN	graph neural network
GUI	graphical user interface
HIV-1	Human immunodeficiency virus type 1
ICS	Interface Contact Similarity score
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
iPAE	interface predicted aligned error
ipLDDT	interface predicted local distance difference test
ipSAE	interaction prediction Score from Aligned Errors
ipTM	interface predicted template modeling score
I-RMSD	interface root mean square deviation
JSON	JavaScript Object Notation
IDDT	local distance difference test
L-RMSD	Ligand Root Mean Square deviation
MC	model confidence
mmCIF	macromolecular Crystallographic Information File
mRNA	messenger RNA
MSA	multiple sequence alignment
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
PAE	predicted aligned error
PCC	Pearson correlation coefficient
PDB	Protein Data Bank
pDockQ	predicted DockQ
pDockQ2	predicted DockQ, extended version for multimeric assemblies
PISA	Proteins, Interfaces, Structures and Assemblies
PLIP	Protein-Ligand Interaction Profiler
pLDDT	predicted local distance difference test
PPI	protein-protein interaction
PSI	Protein Structure Initiative
PSAIA	Protein Structure and Interaction Analyzer
pTM	predicted template modeling score

RDC	residual dipolar coupling
RMSD	root mean square deviation
RNA	ribonucleic acid
ROC	receiver operating characteristic
rRNA	ribosomal RNA
SAbDab	Structural Antibody Database
SASA	solvent-accessible surface area
SCF	Sequence-Coloring Format
TBM	template-based modeling
TM-score	Template Modeling score
TN	true negative
TP	true positive
TPR	true positive rate
tRNA	transfer RNA
UCSF	University of California, San Francisco
VMD	Visual Molecular Dynamics
VoroIF	Voronoi interface score (abbreviated form of VoroIF-GNN)
VoroIF-GNN	Voronoi interface graph neural network score

1

Introduction

Contents

1.1	General background and motivation	2
1.1.1	Macromolecular interactions as the foundation of biology . .	3
1.1.2	The structural basis of biomolecular function	4
1.1.3	Protein-protein and nucleic acid-containing complexes in disease and drug discovery	5
1.2	Protein complex structure determination	5
1.2.1	The structure of proteins	6
1.2.2	Structural principles of protein-protein interfaces	7
1.2.3	Experimental structure determination of proteins and protein complexes	10
1.2.4	Deposition and access of structural data	12
1.2.5	Limitations of X-ray crystallography and cryo-EM	13
1.3	Protein complex structure prediction	14
1.3.1	Computational protein structure prediction	14
1.3.2	Computational prediction of multi-chain complexes	17
1.4	Analysis and evaluation of protein-protein interfaces	18
1.4.1	Molecular visualization of protein structures and complexes .	19
1.4.2	Tools for protein complex interface identification, analysis, and clustering	19
1.4.3	Reference-based quality metrics	20
1.4.4	Scoring metrics for AI-generated protein complex predictions	21
1.4.5	Benchmarking initiatives in structural model evaluation . . .	23
1.4.6	Motivation for PICKLUSTER	23
1.4.7	Motivation for the benchmarking of scoring metrics and C2Qscore	24
1.5	Beyond proteins: Expanding to nucleic acid-containing complexes	25
1.5.1	Biological roles and structural features of nucleic acid-containing complexes	25
1.5.2	Challenges in structural determination and prediction of nucleic acid-containing complexes	28

1.5.3	Computational tools for the analysis of nucleic acid-containing complexes	29
1.5.4	Motivation for ProNA3D	30

Biomolecular interactions constitute the foundation of cellular function, playing essential roles in processes such as signal transduction, transcription, translation, DNA replication and repair, and immune recognition (Liljas et al. 2009; De Las Rivas and Fontanillo 2010; Voet et al. 2016). Structural characterization of these interactions has been central to improving our understanding of these mechanisms (M. C. Thompson et al. 2020). Recent advances in artificial intelligence (AI)-based structure prediction, particularly AlphaFold2 and AlphaFold3, have substantially expanded the availability of structural models for biomolecular complexes across diverse interaction types, including protein-protein, protein-RNA, protein-DNA, and nucleic acid-only assemblies (Jumper et al. 2021; Evans et al. 2022; Abramson et al. 2024; L. Chen et al. 2024). Despite this progress, systematic and confidence-aware approaches for characterizing and assessing predicted interfaces remain limited, particularly across the different biomolecular interaction classes.

This thesis addresses the challenge of characterizing, evaluating, and analyzing the interfaces of biomolecular complexes in both experimentally determined and AI-predicted structures, with particular emphasis on their relevance to structural biology and structure-guided drug discovery. It is structured around three interconnected contributions.

First, PICKLUSTER (Genz et al. 2023) introduces a framework for the identification, spatial subdivision, and confidence-based analysis of protein-protein interfaces. Second, a systematic evaluation of scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions is presented (Genz et al. 2025a), aiming at identifying the most reliable individual metrics. This includes the development of C2Qscore, a combined quality measure designed to provide a more robust assessment of predicted complex quality. Third, ProNA3D (Genz and Topf 2026) extends these concepts to the analysis of protein-nucleic acid and nucleic acid-only complexes. Together, these contributions build on established concepts in structural biology and bioinformatics, including protein and nucleic acid structure, experimental structure determination, and computational structure prediction.

The following chapter outlines the theoretical background of this work. It first reviews the biological significance of biomolecular interactions and their relevance as therapeutic targets, followed by an overview of experimental and computational methods for structure determination and prediction. It then introduces the principles of protein-protein interface analysis, discusses the challenges associated with assessing the quality of predicted complexes, and reviews existing approaches. These concepts are subsequently extended to nucleic acid-containing complexes. The chapter concludes by identifying the specific methodological gaps that motivate the aims of this thesis.

1.1 General background and motivation

This section introduces the major classes of macromolecular interactions and their roles in cellular function, followed by an overview of the structural principles governing molecular recognition and protein organization. The section concludes

by highlighting the relevance of biomolecular interfaces to human disease and structure-guided drug discovery.

1.1.1 Macromolecular interactions as the foundation of biology

Proteins and nucleic acids form dynamic interaction networks that regulate essential cellular processes, including signal transduction, transcription, translation, DNA replication and repair, and immune recognition (Liljas et al. 2009; Voet et al. 2016). These macromolecular interactions determine the specificity and regulation of biochemical pathways (Brown and Sacks 2009; Westermarck et al. 2013). In their absence, cellular function would not be possible.

Protein-protein interactions (PPIs) represent one of the most extensively characterized classes of macromolecular interactions, reflecting the central role of proteins in cellular function (Figure 1.1) (Liljas et al. 2009; Elhabashy et al. 2022). The network of PPIs in humans, referred to as the interactome, is estimated to comprise approximately 650,000 interactions (Stumpf et al. 2008). These interactions exhibit diverse binding modes, ranging from stable, permanent complexes to transient and reversible contacts, particularly in signaling pathways (Elhabashy et al. 2022). In these contexts, proteins function both as signal-transducing effectors and as scaffolds that organize multi-protein assemblies (Liljas et al. 2009; Good et al. 2011). PPIs regulate key biological processes such as receptor activation, cellular architecture and organization, enzymatic catalysis, post-translational modification, and energy transformation (Branden and Tooze 1999; Liljas et al. 2009). Disruptions of specific PPIs are associated with numerous human diseases, including cancer, neurodegenerative disorders, genetic disorders, and infectious diseases (Bergendahl et al. 2019; F. Cheng et al. 2021; W. Zhang et al. 2023; Jia et al. 2024).

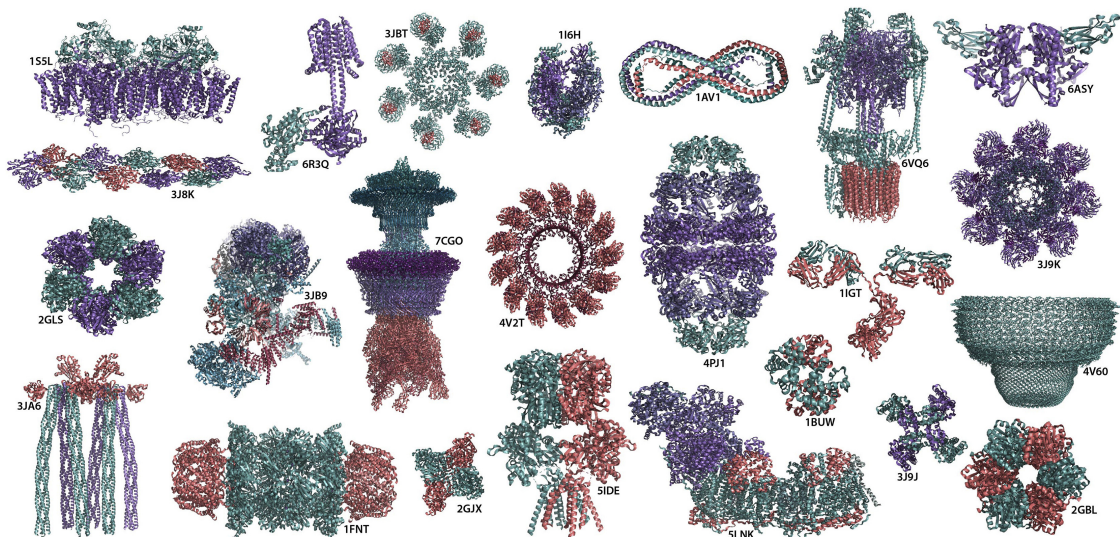


Figure 1.1: A collection of macromolecular complexes formed by protein-protein interactions. Representative examples of structurally diverse protein assemblies, illustrating the range of sizes, stoichiometries, and biological functions mediated by protein-protein interfaces. The corresponding PDB IDs are indicated below each structure. Adapted from Elhabashy et al. (2022)

In addition to protein-protein interactions, protein-nucleic acid interactions also play a central role in cellular function. Protein-DNA interactions are involved in processes

such as transcription, chromatin remodeling, and DNA replication and repair, while protein-RNA interactions regulate RNA processing, splicing, transport, translation, stability, and ribonucleoprotein assembly (Hertel and Graveley 2005; Noller 2005; Ptashne 2005; Liljas et al. 2009; Hudson and Ortlund 2014). Beyond these, nucleic acid-only interactions further contribute to cellular processes. DNA-DNA, RNA-RNA, and RNA-DNA complexes not only encode and transmit genetic information but also perform essential structural and catalytic functions. Examples include RNA-RNA interactions in ribozymes and ribosomes, RNA-DNA hybrids that serve as intermediates in replication and transcription, and DNA-DNA complexes involved in homologous recombination and higher-order chromosome organization (Branden and Tooze 1999; Liljas et al. 2009; Alonso and Mondragón 2021). The structural diversity of these interaction types reflects the broad range of biological functions they support and highlights the need for analytical frameworks capable of addressing the full spectrum of macromolecular assemblies.

1.1.2 The structural basis of biomolecular function

The three-dimensional arrangement of atoms within a macromolecule is the primary determinant of its biological function (Liljas et al. 2009). Molecular recognition, defined as the selective and specific interaction between macromolecules, arises from the geometric and physicochemical complementarity of their contact surfaces (Chothia and Janin 1975; Janin 1995). A key feature is the shape complementarity, which is closely associated with van-der-Waals interactions and non-polar contacts (Lawrence and Colman 1993; Q. Zhang et al. 2009; Y. Li et al. 2013). Electrostatic interactions and non-covalent interactions, such as hydrogen bonds, determine the affinity and specificity of the interaction (Chothia and Janin 1975; Liljas et al. 2009; Adhav and Saikrishnan 2023). These principles are conserved across all classes of macromolecular interactions, including protein-protein, protein-nucleic acid, and nucleic acid-only complexes (Liljas et al. 2009; Anashkina 2023). Consequently, the specificity of these interactions is determined by their three-dimensional structure rather than sequence alone.

The determination of the three-dimensional structures of myoglobin and hemoglobin by X-ray crystallography established the link between the three-dimensional structure and the function of a protein (Kendrew et al. 1960; Perutz et al. 1960; Blundell 2021). These and subsequent studies demonstrated that the spatial arrangement of atoms within a protein determines its ability to bind ligands, undergo conformational transitions, and catalyze chemical reactions. This established the structure-function paradigm as one of the key concepts of molecular biology, which has since been widely confirmed (Branden and Tooze 1999; Liljas et al. 2009; M. C. Thompson et al. 2020; Blundell 2021).

Macromolecular function, however, cannot be fully captured by a single static conformation. Proteins and nucleic acids are inherently dynamic, and functionally relevant structural changes, including domain rearrangements, allosteric transitions, and induced-fit binding, are crucial to their biological roles (Fenwick et al. 2011; Guo and H.-X. Zhou 2016). As a result, a single structure often represents only one of several existing interaction modes, which has important consequences for interface analysis (Srivastava et al. 2018).

Structural information is therefore essential for understanding macromolecular interactions. Detailed knowledge of interface geometry and physicochemical properties provides the basis for understanding biological specificity and for the rational design

of molecules that modulate these interactions. The methods used to obtain such information are described in the following sections.

1.1.3 Protein-protein and nucleic acid-containing complexes in disease and drug discovery

Given their central role in cellular function, dysregulated protein-protein interactions underlie a wide range of human diseases (Ryan and Matthews 2005; F. Cheng et al. 2021). Oncogenic signaling complexes, viral capsid assemblies, and pathological protein aggregates all depend on aberrant or dysregulated protein-protein contacts, identifying these interfaces as attractive targets for therapeutic intervention (Nero et al. 2014; F. Cheng et al. 2021; W. Zhang et al. 2023; Jia et al. 2024). Over the past decades, successful efforts to inhibit PPIs using small molecules have demonstrated their suitability as drug targets, leading to the development of several U.S. Food and Drug Administration (FDA)-approved therapeutics, including the small-molecule inhibitors venetoclax (BCL-2), maraviroc (CCR5), sotorasib (RAS GTPase family), and adagrasib (KRAS G12C mutant) (Lieberman-Blum et al. 2008; Deeks 2016; Blair 2021; Dhillon 2023; Nada et al. 2024).

Protein-nucleic acid interfaces constitute an equally important class of therapeutic targets. Disruption of protein-nucleic acid interactions can result in transcriptional misregulation, defective RNA processing, and impaired genome maintenance (Branden and Tooze 1999; Hakem 2008; Järvelin et al. 2016). Protein-RNA interactions are frequently altered in a variety of diseases, including viral infections that rely on RNA-binding proteins and polymerases for genome replication and packaging (Tao and Q. Ye 2010; Z. Li and Nagy 2011; Genz et al. 2025b). In oncology, dysregulation of RNA-binding proteins and splicing factors leads to aberrant RNA processing with pathological consequences such as tumor progression and resistance to apoptosis (Kang et al. 2020; Qin et al. 2020). Similarly, protein-DNA interfaces, including those involving oncogenic transcription factors such as MYC and STAT3, as well as DNA repair enzymes such as BRCA1, are implicated in genome instability and defective repair mechanisms (Look 1997; N. Tuteja and R. Tuteja 2001; Darnell 2002; Genz et al. 2025b).

The exploitation of these interfaces as therapeutic targets requires detailed structural knowledge of their geometry and physicochemical properties, thereby motivating the development and application of experimental methods for structure determination.

1.2 Protein complex structure determination

Interpreting the function and interactions of protein complexes requires knowledge of their three-dimensional structure at atomic or near-atomic resolution. This section outlines the structural organization of proteins and protein-protein interfaces, introduces the experimental methods used for structure determination, and discusses the limitations of these approaches that motivate the development and application of computational prediction methods described in the following section.

1.2.1 The structure of proteins

Proteins are biological macromolecules composed of amino acid residues linked by peptide bonds into a polypeptide chain. Protein structure is conventionally described at four hierarchical levels (Figure 1.2) (Branden and Tooze 1999; Liljas et al. 2009). The primary structure corresponds to the linear sequence of amino acid residues, which encodes the information required to specify higher-order structural organization. The secondary structure describes recurring local conformations of the polypeptide backbone, stabilized by hydrogen bonds between backbone amide and carbonyl groups. The most common elements are α -helices and β -sheets, as well as connecting loops and coil regions. The tertiary structure refers to the overall three-dimensional fold of a single polypeptide chain, determined primarily by interactions between amino acid side chains, including hydrophobic contacts, hydrogen bonds, electrostatic interactions, and disulfide bridges, in combination with backbone geometry constraints. Many proteins are organized into distinct folding units termed domains, which constitute the modular building blocks of the overall tertiary structure. Proteins consisting of a single polypeptide chain are termed monomeric. In contrast, when multiple polypeptide chains associate, the resulting assembly is described as the quaternary structure, which defines the spatial arrangement of subunits and the interactions that stabilize the complex. These intermolecular contacts are fundamental to protein function in the cellular context.

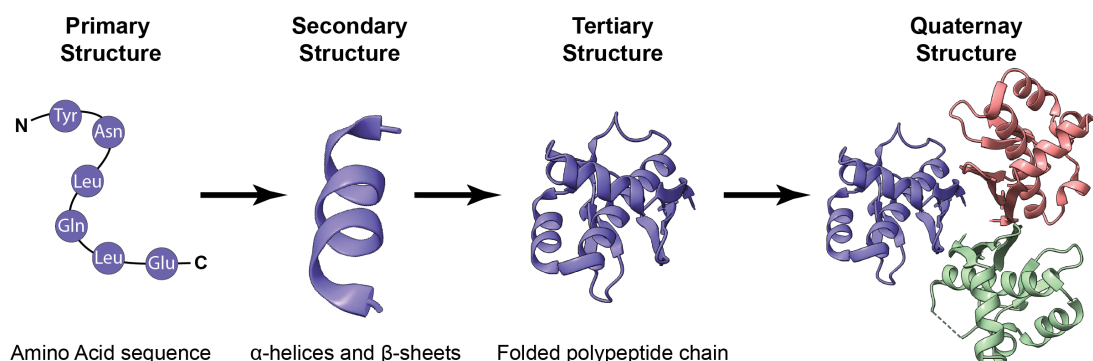


Figure 1.2: The four hierarchical levels of protein structure. From left to right: primary structure (amino acid sequence), secondary structure (local backbone conformations, including α -helices and β -sheets), tertiary structure (three-dimensional fold of a single polypeptide chain and the spatial arrangements of domains), and quaternary structure (spatial arrangement of multiple chains in a multi-subunit complex), exemplified by the *Streptococcus pneumoniae* acyl carrier protein synthase (PDB ID: 1FTH).

The chemical diversity of proteins arises from the side chains of 20 standard L-amino acids, which differ in size, charge, polarity, and capacity for specific non-covalent interactions, and is further expanded by a wide range of non-standard and post-translationally modified residues (Liljas et al. 2009). The side chains of the standard amino acids are commonly classified into three groups: nonpolar (hydrophobic) residues (alanine, valine, leucine, isoleucine, methionine, proline, and phenylalanine), polar uncharged residues (serine, threonine, asparagine, glutamine, cysteine, tyrosine, and tryptophan), and charged residues, which are further subdivided into positively charged (lysine, arginine, histidine) and negatively charged (aspartate, glutamate) amino acids (Branden and Tooze 1999; Liljas et al. 2009). Glycine, possessing only a hydrogen atom as its side chain, exhibits exceptional conformational flexibility.

Except for proline, all amino acids share a common backbone structure consisting of a central C_α atom bonded to a hydrogen atom, a primary amino group, and a carboxyl group (Branden and Tooze 1999). In proline, the side chain forms a covalent bond with the backbone nitrogen, creating a cyclic secondary amine that restricts the backbone dihedral angle and prevents standard hydrogen bond donation, making it a common disruptor of regular secondary structure elements (Liljas et al. 2009). Hydrophobic residues are typically buried within the protein core, whereas polar and charged residues are enriched at the surface, generating the heterogeneous surface properties that govern molecular recognition (Chothia and Janin 1975; Jones and Thornton 1996; C. Yan et al. 2008; Liljas et al. 2009). Consequently, the composition and physicochemical properties of residues at an interface are key determinants of binding specificity, affinity, and the suitability of a region for therapeutic targeting (Chothia and Janin 1975; Liljas et al. 2009; London et al. 2013).

Proteins are frequently organized into distinct structural domains, which often perform specific functions or mediate particular interactions (Liljas et al. 2009). In multi-domain proteins, these domains are typically connected by linker sequences that help maintain inter-domain organization (Branden and Tooze 1999; George and Heringa 2002). Linkers may be flexible, enabling conformational rearrangements and inter-domain interactions, or relatively rigid, restricting motions (Reddy Chichili et al. 2013). This modular organization enables proteins to engage multiple interaction partners and results in interfaces that may consist of distinct sub-regions (Keskin et al. 2005).

Not all proteins adopt stable, well-defined three-dimensional structures under physiological conditions. Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) lack a fixed tertiary structure in isolation (Voet et al. 2016). These regions frequently mediate transient protein-protein interactions, particularly in signaling and transcriptional regulation (Perkins et al. 2010; Voet et al. 2016). Their high conformational dynamics and flexibility pose challenges for structural characterization, as disordered regions are often unresolved in experimental structures (Le Gall et al. 2007).

1.2.2 Structural principles of protein-protein interfaces

As described above, protein chains can associate with one or more additional chains to form protein complexes (Chothia and Janin 1975). The regions mediating these interactions are referred to as protein-protein interfaces (Chothia and Janin 1975; De Las Rivas and Fontanillo 2010).

Protein-protein interfaces are characterized by distinct physicochemical and geometric properties (Ponstingl et al. 2005). Hydrophobic interactions constitute the primary thermodynamic driving force for complex stability, whereas non-covalent interactions predominantly contribute to binding specificity through requirements for geometric and chemical complementarity (Chothia and Janin 1975). These interactions include hydrogen bonds, ionic interactions (salt bridges), van der Waals interactions (Figure 1.3), and hydrophobic contacts, as well as weaker interactions such as $C-H \cdots \pi$ contacts and sulfur-mediated hydrogen bonds (Černý and Hobza 2007; Adhav and Saikrishnan 2023). Electrostatic interactions additionally influence both the kinetics of complex formation and binding affinity (Sheinerman et al. 2000; Heifetz et al. 2002). The overall thermodynamic stability of a complex is quantified by its binding free energy (ΔG) and dissociation constant (K_d), which for protein-protein interactions

typically spans the nanomolar to micromolar range (Branden and Tooze 1999; Nooren and Thornton 2003b; Perkins et al. 2010). The free energy of binding is determined by the relationship between enthalpic and entropic contributions according to

$$\Delta G = \Delta H - T\Delta S \quad (1.1)$$

where ΔH is the enthalpic contribution arising from interactions such as hydrogen bonding, electrostatic interactions, and van der Waals forces, ΔS is the entropic contribution associated with desolvation and conformational changes, and T is the absolute temperature in Kelvin.

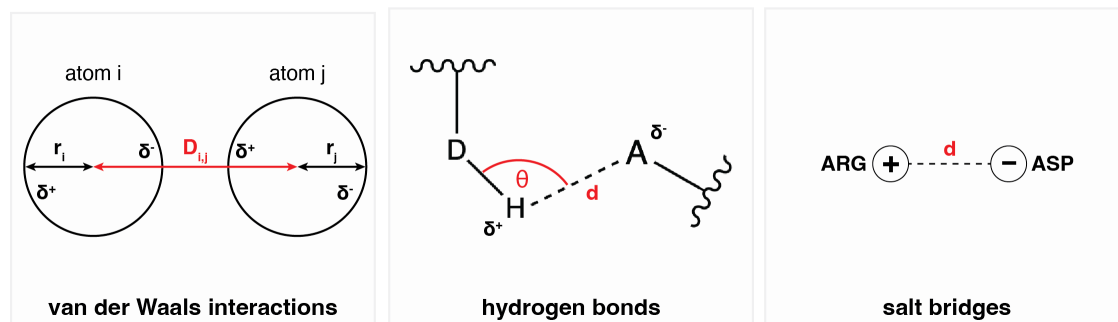


Figure 1.3: Schematic representation of the major non-covalent interactions in protein structures. d : interaction distance, θ : interaction angle, r_i : van der Waals radius of atom i , r_j : van der Waals radius of atom j , $D_{i,j}$: distance between atoms i and j , δ^+ : partial positive charge, δ^- : partial negative charge, $+$: positive charge, $-$: negative charge.

Protein-protein interfaces can be further classified based on the lifetime of the complex (Nooren and Thornton 2003a). While most protein complexes are transient, they exhibit a wide range of affinities and lifetimes (Janin et al. 2008). Transient interfaces typically mediate regulatory or signaling interactions and are characterized by relatively weak binding affinities and short lifetimes, ranging from seconds to minutes, as exemplified by interactions involving p53 (Perkins et al. 2010). In contrast, obligate interfaces occur in proteins that do not exist as stable monomers in isolation and are permanently present within assembled complexes. These interfaces exhibit stronger binding affinities and much longer lifetimes, often remaining associated for hours and, in the case of large molecular machines such as the ribosome and the proteasome, for the entire cellular lifespan of the assembly (Nooren and Thornton 2003a; Mintseris and Weng 2005).

A further characteristic of protein interfaces is their size. A commonly used measure is the buried surface area (BSA), defined as the difference in solvent-accessible surface area (SASA) between the unbound monomers and the assembled complex (Janin et al. 1988; Janin et al. 2008). Analyses of high-resolution crystal structures of protein complexes deposited in the Protein Data Bank (PDB) (Berman et al. 2000) have shown that BSA values span a wide range, from below 1,000 Å² to several tens of thousands of Å², reflecting the diversity of protein-protein interactions (Janin et al. 1988; Janin et al. 2008). The burial of nonpolar surface area provides a major thermodynamic driving force for complex formation (Keskin et al. 2005).

Protein complexes composed of identical polypeptide chains are termed homomeric, whereas those consisting of different chains are referred to as heteromeric (Nooren and Thornton 2003a; Ofra and Rost 2003; C. Yan et al. 2008). These classes differ

in their structural properties, as homomeric interfaces are, on average, larger, more hydrophobic, and more evolutionarily conserved than heteromeric interfaces (C. Yan et al. 2008). More generally, interface residues tend to be more conserved than other surface-exposed residues, reflecting the functional importance of maintaining compatible binding surfaces (Yao et al. 2003).

Protein-protein interfaces are often described as relatively large and geometrically flat, which historically led to the perception that they were challenging targets for small-molecule binding. (Chothia and Janin 1975; Jones and Thornton 1996; Wells and McClendon 2007). However, several studies have shown that binding free energy is distributed unevenly across the interface (Clackson and Wells 1995; Bogan and Thorn 1998; Wells and McClendon 2007; London et al. 2013; H. Jubb et al. 2015). A small subset of residues, termed hotspots, contributes the majority of the binding interaction energy and frequently involves large side chains containing amino acids such as arginine and tyrosine (Figure 1.4). Experimental approaches such as alanine-scanning mutagenesis enable the identification of these energetically important residues (Cunningham and Wells 1991; Keskin et al. 2005). The success of small-molecule inhibitors targeting protein-protein interactions described in subsection 1.1.3 has highlighted these interfaces as promising therapeutic targets. Notably, hotspot residues tend to cluster within spatially localized sub-regions of the interface, suggesting that the identification and characterization of such sub-interfaces may provide a promising strategy for structure-guided drug discovery.

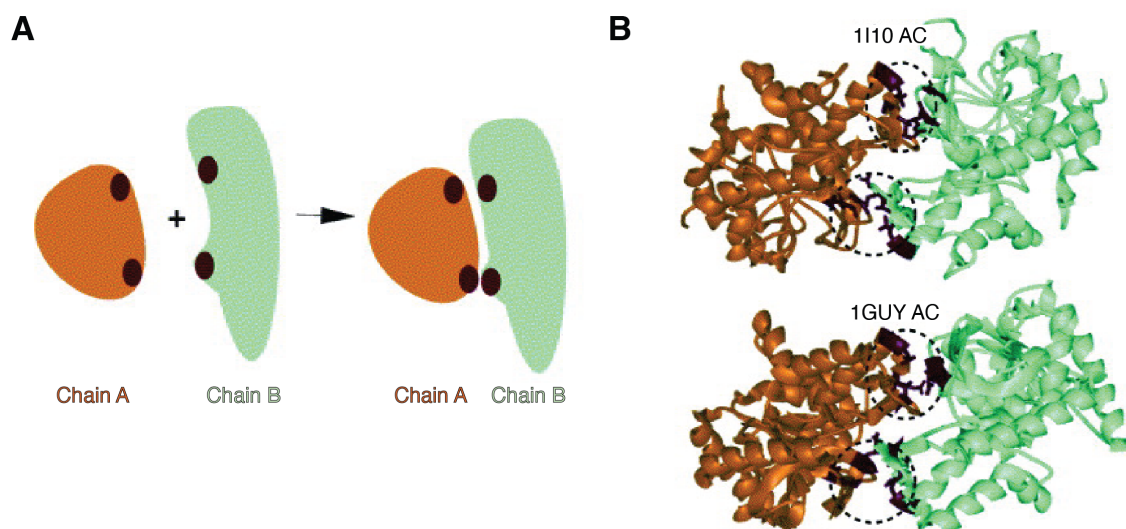


Figure 1.4: The organization of hotspots within protein-protein interfaces. (A) Schematic representation of two proteins before and after binding, with hotspot clusters highlighted in black, demonstrating their interaction across the interface. (B) Ribbon diagrams of two oxidoreductase proteins, with hotspot residues shown in purple. These residues are spatially clustered rather than randomly distributed, forming distinct regions within the interface (circled in black). Adapted from Keskin et al. (2005).

1.2.3 Experimental structure determination of proteins and protein complexes

The first high-resolution protein structures, those of myoglobin (PDB ID: 1MBN, resolution: 2.0 Å) and hemoglobin (PDB ID: 2DHB, resolution: 2.8 Å), were determined by X-ray crystallography (Kendrew et al. 1960; Perutz et al. 1960). This technique is well established for the structural determination of biological macromolecules in crystalline form and remains one of the most widely used methods in structural biology (Figure 1.5) (Tripathy 2025; Rosenhouse-Dantsker 2026). X-ray crystallography requires the preparation of well-ordered crystals of the purified target molecules, which are exposed to an X-ray beam to produce a characteristic diffraction pattern (Tripathy 2025). The diffraction pattern provides structure factor amplitudes, which, together with estimated phases, are used in a Fourier transform to calculate the electron density map from which the atomic coordinates of the target molecule are derived (Rhodes 2010).

Resolution in X-ray crystallography, reported in Å, describes the smallest distance that can be distinguished in the diffraction data and resulting electron-density map, with lower values corresponding to higher structural detail (Branden and Tooze 1999; Smyth and Martin 2000; Dubach and Guskov 2020). To identify overall structural features, a resolution of at least 3.0 Å is required. A typical carbon-carbon bond length is approximately 1.5 Å, and at resolutions higher than 3 Å individual side chains can often be resolved in the electron density (Branden and Tooze 1999; Smyth and Martin 2000). A related parameter is the crystallographic B-factor, also referred to as the atomic displacement parameter, which quantifies the mean-square displacement of the electron density of an atom from its average position (Trueblood et al. 1996; Rhodes 2010). B-factors reflect both thermal motion and static disorder within the structure. Elevated B-factor values are associated with increased positional uncertainty and are commonly observed in flexible loops, disordered regions, and solvent-exposed surface areas (Radivojac et al. 2004; Afonine et al. 2012). X-ray crystallography does not inherently impose a strict upper limit on the size of the macromolecular assembly. Very large complexes, including ribosomes, have been successfully crystallized and structurally resolved. However, the method is limited by the requirement to obtain well-ordered, diffracting crystals, which becomes increasingly challenging for large, flexible, or heterogeneous systems (Grey and D. H. Thompson 2010; Nurullina et al. 2025).

Another approach for obtaining structural information of protein complexes is cryo-electron microscopy (cryo-EM) (Figure 1.5). Over the past decade, cryo-EM has undergone a fundamental transformation and has emerged as a powerful technique for resolving large macromolecular assemblies at near-atomic resolution without the need for crystallization, thereby enabling structure determination under near-native conditions (Dubochet et al. 1988; Namba and Makino 2022; Tripathy 2025; Rosenhouse-Dantsker 2026). In single-particle cryo-EM, purified macromolecules are vitrified in a thin layer of amorphous ice and imaged with a transmission electron microscope under cryogenic conditions, which preserves their native-like conformations (Frank 2002; Bai et al. 2013; Voet et al. 2016; Henderson 2018; Rosenhouse-Dantsker 2026). Subsequently, large numbers of particle images are computationally aligned and averaged to reconstruct a three-dimensional density map, considering conformational variability and sample heterogeneity (Y. Cheng et al. 2015). Recent advances in detector technology and image processing have enabled resolutions approaching 1.25 Å, making cryo-EM competitive with X-ray crystallography (Y. Cheng 2015;

Mendez et al. 2019; Yip et al. 2020; Beckers et al. 2021). In cryo-EM, resolution reflects the level of structural detail present in the reconstructed density map. It is conventionally estimated using the Fourier Shell Correlation (FSC) between two independently reconstructed half-maps, with the resolution defined as the spatial frequency at which the FSC falls to 0.143, known as the gold-standard criterion (Rosenthal and Henderson 2003; Heel and Schatz 2005).

Unlike X-ray crystallography, however, the resolution is not uniform across the map but varies locally (Dubach and Guskov 2020). Therefore, a single global resolution value provides only an overall estimate of map quality. Factors such as conformational heterogeneity, molecular flexibility, radiation damage, and limitations in image processing can lead to differences in local resolution, necessitating the use of local resolution estimation methods (Cardone et al. 2013; Dubach and Guskov 2020). Cryo-EM is particularly well-suited for large macromolecular assemblies, such as ribosomes, viral capsids, and membrane protein complexes, and it is not constrained by an upper size limit. However, the structural determination of small macromolecules (typically below ~100 kDa) remains challenging due to their low contrast in vitreous ice (Wentinck et al. 2022). The ability to capture macromolecules in near-native states and to resolve multiple conformations has made cryo-EM a key method in modern structural biology.

Nuclear magnetic resonance (NMR) spectroscopy is a complementary method for determining structures in solution under near-physiological conditions, enabling the study of dynamic proteins and transient interactions (Figure 1.5) (Voet et al. 2016; Rosenhouse-Dantsker 2026). The method typically requires isotopic labeling and is most effective for proteins below approximately 50 kDa (Krishnan and Rupp 2012; Rosenhouse-Dantsker 2026). Despite this size restriction, NMR can provide insight into protein-protein interactions through residual dipolar couplings (RDCs), which yield orientational restraints that characterize the spatial arrangement of binding partners (Marion 2013). However, NMR generally provides lower structural resolution than X-ray crystallography and cryo-EM, and spectral assignment and data interpretation remain labor-intensive (Krishnan and Rupp 2012; Rosenhouse-Dantsker 2026). Since the structural datasets analyzed in this thesis are predominantly derived from X-ray crystallography and cryo-EM, NMR spectroscopy is not discussed in further detail.

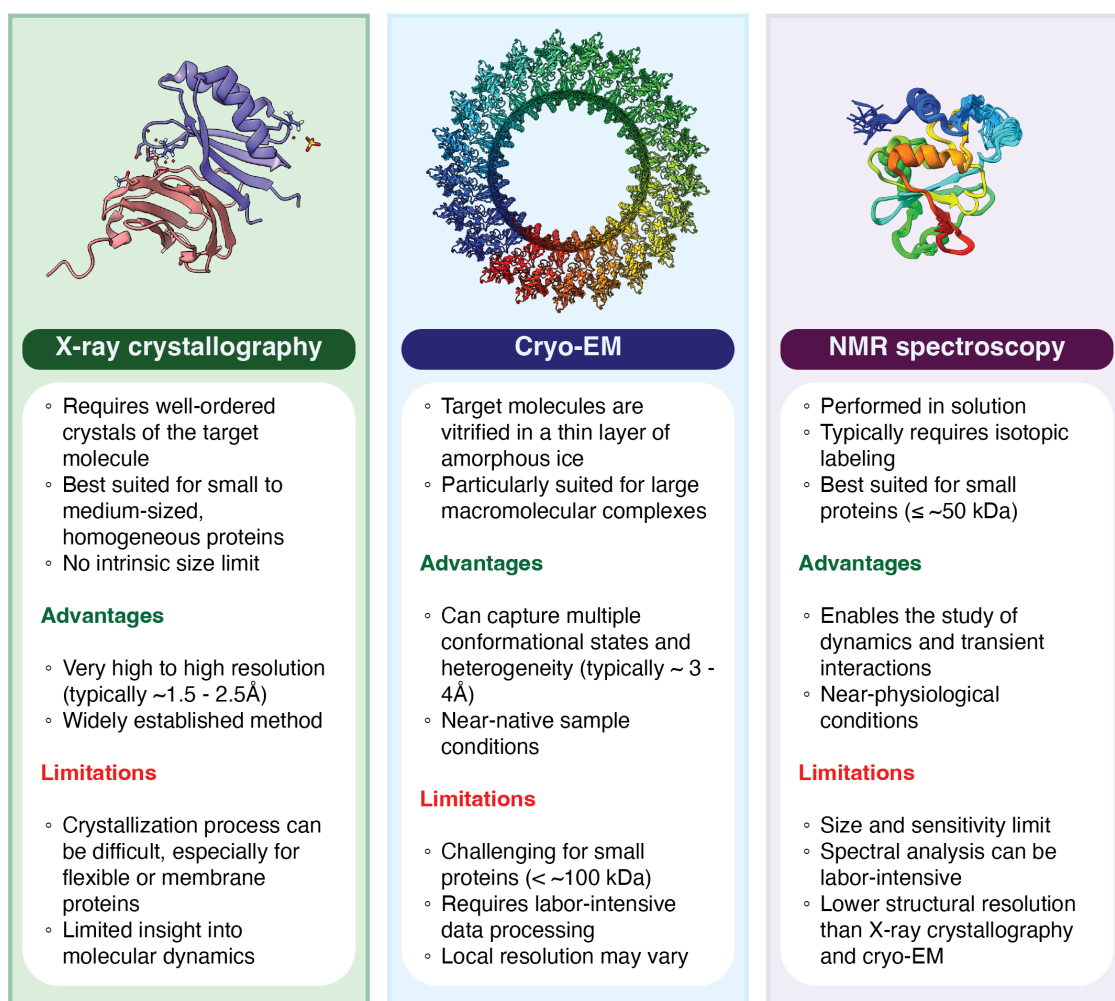


Figure 1.5: Comparison of X-ray crystallography, cryo-EM, and NMR spectroscopy, highlighting their experimental principles, methodological advantages, and limitations. Representative structures are shown for each technique (PDB IDs: 8BR2, 8B6W, 9BU5).

1.2.4 Deposition and access of structural data

The Protein Data Bank (PDB) (<https://www.rcsb.org/>) is the global repository for experimentally determined three-dimensional structures of biological macromolecules, including proteins, nucleic acids, and their complexes (Berman et al. 2000; Burley et al. 2017). Since its establishment in 1971, it has grown to contain over 250,000 entries as of 2026, obtained by X-ray crystallography, NMR spectroscopy, and cryo-EM (*PDB Statistics* 2026). Each entry includes validated atomic coordinates, experimental data, and metadata. The PDB exhibits compositional biases, with a predominance of X-ray structures of soluble proteins or isolated domains, while membrane proteins, intrinsically disordered proteins, and large multi-component assemblies remain underrepresented relative to their biological prevalence (Le Gall et al. 2007). Structures containing nucleic acids, including protein-RNA, protein-DNA, and nucleic acid-only complexes, constitute a comparatively small fraction of the database, reflecting the experimental challenges associated with their structural characterization (Figure 1.6).

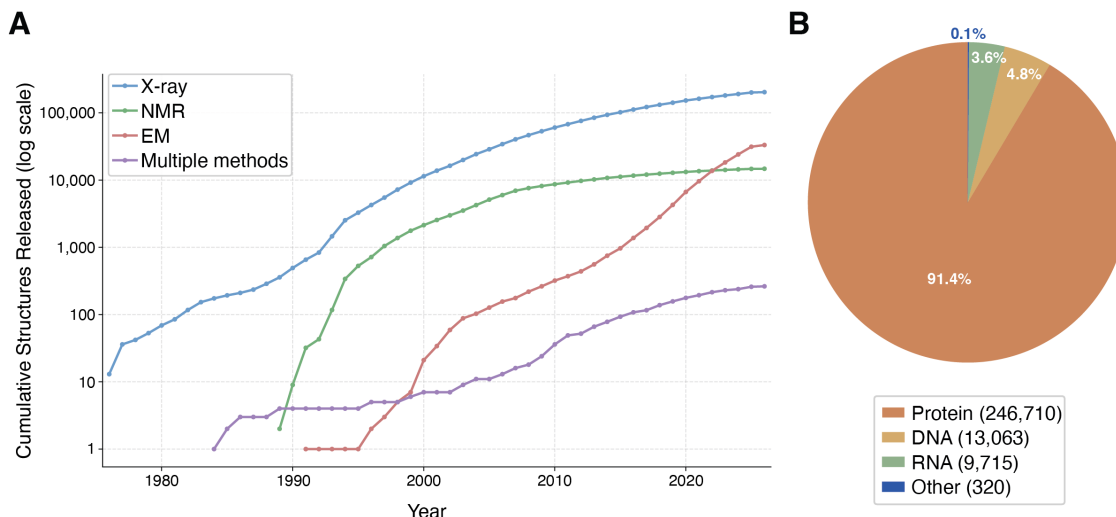


Figure 1.6: Growth and composition of the Protein Data Bank. (A) Cumulative number of deposited structures per year (log scale), grouped by experimental method: X-ray crystallography (blue), NMR spectroscopy (green), electron microscopy (red), and multi-method structures (purple). (B) Composition of all PDB entries by molecule type (PDB accessed 07.04.2026).

1.2.5 Limitations of X-ray crystallography and cryo-EM

Both X-ray crystallography and cryo-EM require the production of sufficient quantities of pure and homogeneous target protein. This can be technically challenging, particularly for membrane proteins, multi-subunit assemblies, glycosylated proteins, and proteins that are unstable in isolation (Smyth and Martin 2000; Namba and Makino 2022; Tripathy 2025).

For X-ray crystallography, the growth of well-diffracting crystals is typically the major bottleneck, requiring extensive screening of crystallization conditions and large quantities of highly concentrated protein (Grey and D. H. Thompson 2010). Crystallization is often unsuccessful for proteins with flexible regions, transmembrane domains, or extensive post-translational modifications (Smyth and Martin 2000; Srivastava et al. 2018). Additionally, crystal packing can impose conformational constraints and stabilize binding states that may not reflect physiological states (Schweke et al. 2023). Both X-ray and electron beams cause radiation damage, which progressively degrades data quality (Y. Cheng 2015). Furthermore, flexible and intrinsically disordered regions are often poorly resolved or absent in both X-ray and cryo-EM structures, as their conformational heterogeneity prevents the formation of well-defined electron density or stable cryo-EM reconstructions (Le Gall et al. 2007).

Cryo-EM presents specific limitations related to conformational heterogeneity and particle behavior on the grid. In single-particle datasets, structural heterogeneity arises from intrinsic protein dynamics (Rosenhouse-Dantsker 2026). Discrete conformational states can often be separated by 3D classification, whereas continuous conformational variability is more difficult to resolve and can lead to blurred density and reduced local resolution (Tripathy 2025; Rosenhouse-Dantsker 2026). Approaches such as 3D variability analysis can partially capture this heterogeneity, but are limited by data quality, particle number, and the complexity of the conformational landscape (D’Imprima and Kühlbrandt 2021; Rosenhouse-Dantsker 2026). Additionally, preferred particle orientations towards the air-water interface in the vitreous ice can lead to

incomplete angular sampling, negatively affecting the quality of the reconstructed map (Y. Cheng et al. 2015).

Both X-ray crystallography and cryo-EM present significant practical limits. Despite advances in automation and data processing, the acquisition and analysis of diffraction data and cryo-EM datasets, as well as sample preparation and optimization, remain technically demanding, time-intensive processes that require specialized training and expertise (Y. Cheng 2015). In addition, the installation and operation of the necessary infrastructure, including high-end microscopes or diffractometers, detectors, and computational resources, require substantial financial investment (Y. Cheng 2015; Rosenhouse-Dantsker 2026).

To address these limitations, large-scale structural genomics initiatives were established in the early 2000s to enable systematic high-throughput protein structure determination (Burley 2000). Programs such as the Protein Structure Initiative (PSI) introduced automated and standardized workflows to increase structural coverage across protein families (Burley 2000; Montelione 2012). Despite substantially accelerating structure deposition, experimental bottlenecks, including sample preparation, crystallization, and data processing, remained major challenges, highlighting the need for complementary computational methods.

1.3 Protein complex structure prediction

The limitations of experimental structure determination have motivated the development of computational approaches that enable structural modeling *in silico* without the need for physical samples or specialized instrumentation. This section outlines the development of computational protein structure prediction, from template-based methods to deep learning approaches such as AlphaFold2 and AlphaFold3, and discusses the extension of these approaches to multi-chain complexes.

1.3.1 Computational protein structure prediction

The prediction of the three-dimensional fold of a protein structure from an amino acid sequence alone has been one of the central challenges of computational structural biology for decades. The Critical Assessment of Structure Prediction (CASP), a biennial community-wide blind prediction experiment initiated in 1994, has provided a standardized framework for evaluating progress in the field by comparing submitted models against unpublished experimental structures (Moult et al. 1995; Kryshtafovych et al. 2019, 2026).

Comparative modeling, also referred to as homology modeling, was the first broadly successful computational approach to protein structure prediction. As a template-based modeling (TBM) method, it is based on the assumption that proteins with similar sequences tend to adopt similar three-dimensional structures (Martí-Renom et al. 2000; Rosenhouse-Dantsker 2026). The workflow typically involves identifying suitable template structures in the PDB, aligning the target sequence to one or more templates using a multiple sequence alignment (MSA), and constructing a three-dimensional model by transferring conserved structural features (Figure 1.7) (Moult et al. 1995; Martí-Renom et al. 2000). First, the backbone is modeled, followed by variable regions such as loops and the addition of side chains, which are essential for predicting protein

function and interactions (Rosenhouse-Dantsker 2026). The final model is then evaluated for quality (Martí-Renom et al. 2000). Among the most widely used implementations of this approach is MODELLER (Sali and Blundell 1993), which generates structures by satisfying spatial restraints derived from the template alignment. While highly effective when close homologs of the target are available, the accuracy of homology modeling declines as sequence identity decreases, and the method cannot be applied in the absence of suitable templates (Rosenhouse-Dantsker 2026).



Figure 1.7: Schematic overview of the homology modeling workflow. The process includes template identification, multiple sequence alignment (MSA), template-based model generation, and structural evaluation and validation.

In such cases, *de novo* methods provide an alternative. These template-free modeling (FM) approaches aim to identify the most stable protein conformation through conformational sampling guided by physics-based or chemistry-based scoring functions, without relying on homology (Thachuk et al. 2007; Rosenhouse-Dantsker 2026). They are based on the assumption that the native conformation of a protein corresponds to a minimum free energy state (Rosenhouse-Dantsker 2026). Consequently, they require accurate estimation of conformational free energies and efficient exploration of the conformational space (Xu and Y. Zhang 2012). Approaches such as Rosetta (Rohl et al. 2004) generate models by assembling fragments from known protein folds and evaluating them with empirical energy functions. While these approaches achieved substantial success for short proteins, their accuracy for longer sequences remained limited (Rosenhouse-Dantsker 2026).

The integration of co-evolutionary information from MSAs represented a key advance in protein structure prediction (Rosenhouse-Dantsker 2026). Residue pairs that are in spatial proximity in the folded structure are constrained by evolution, enabling the prediction of residue contacts (Göbel et al. 1994; Rosenhouse-Dantsker 2026). These contacts are typically represented as binary contact maps derived from alignment statistics and have been shown to significantly improve *de novo* modeling, particularly for proteins lacking close homologs (Kryshtafovych et al. 2019; Rosenhouse-Dantsker 2026).

Building on these developments, end-to-end deep learning methods rapidly transformed protein structure prediction. These approaches extend binary contact maps by predicting distance maps and distance distributions, providing more detailed information about spatial relationships between residue pairs. Additional features such as hydrogen bonds, torsion angles, and orientation-dependent interactions were also incorporated (Rosenhouse-Dantsker 2026). At CASP13 in 2018 (Figure 1.9), AlphaFold1 (Senior et al. 2020) used deep residual networks trained on co-evolutionary features from MSAs and substantially outperformed other methods in the free modeling

category, establishing deep learning as a leading approach for protein structure prediction (Kryshtafovych et al. 2019; Senior et al. 2019).

At CASP14 in 2020, the end-to-end deep learning method AlphaFold2 achieved near-experimental accuracy across a wide range of monomeric target proteins, marking a breakthrough widely regarded as solving the protein folding problem for most naturally occurring sequences (Figure 1.9) (Jumper et al. 2021; Kryshtafovych et al. 2021). This achievement was further recognized by the award of the Nobel Prize in Chemistry in 2024 (The Nobel Prize 2024).

AlphaFold2 is based on an integrated multi-step architecture (Figure 1.8A) (Jumper et al. 2021; Rosenhouse-Dantsker 2026). It combines feature extraction, distance prediction, and structure generation within a single neural network. The model employs a transformer-based architecture known as the Evoformer to process MSAs and pairwise residue representations, capturing co-evolutionary and structural relationships (Jumper et al. 2021). MSAs are obtained using the tools JackHMMER (Johnson et al. 2010) and HHBlits (Remmert et al. 2012). The final structure module produces three-dimensional coordinates of the protein atoms. Although AlphaFold2 represented a major breakthrough in monomeric protein structure prediction, its architecture was not specifically designed for modeling multi-chain protein complexes.

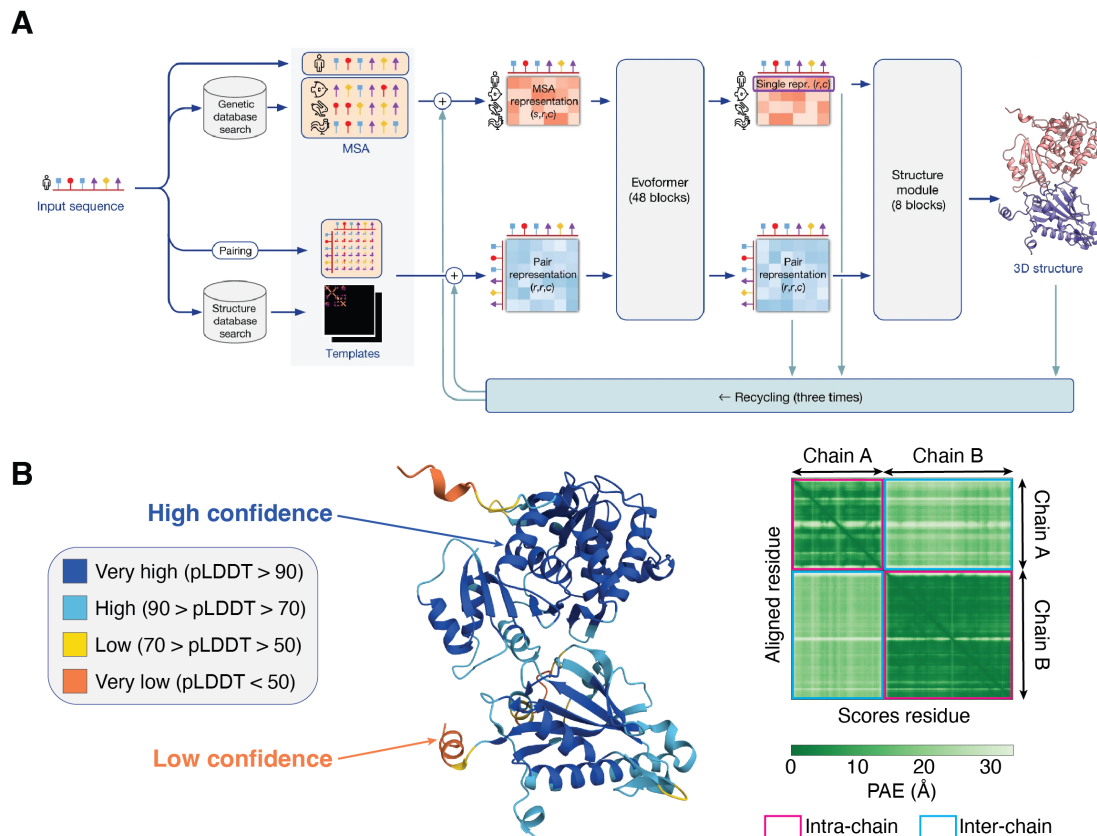


Figure 1.8: Overview of the AlphaFold2 architecture and prediction confidence metrics. (A) Schematic representation of the AlphaFold2 pipeline. Adapted from Jumper et al. (2021). (B) Example pLDDT and predicted aligned error (PAE) confidence metrics for the AlphaFold-Multimer prediction of PDB ID: 7LY5.

1.3.2 Computational prediction of multi-chain complexes

Prior to end-to-end deep learning approaches, the computational prediction of protein complex structures relied primarily on protein-protein docking methods (Ritchie 2008; P. Lin et al. 2024). These approaches use experimentally or computationally determined structures of individual protein chains as input and search for the most favorable binding geometry through rigid-body or flexible sampling of relative orientations and positions (Wodak and Janin 1978; Bonvin 2006; Y. Yan et al. 2020; Barhoon and Mahdiuni 2025). Binding poses are subsequently ranked using energy-, knowledge-, and shape-based scoring functions (Barhoon and Mahdiuni 2025). Depending on the availability of prior experimental information, docking can be performed either *ab initio* or guided by template information and biochemical restraints to reduce the conformational search space. Widely used docking methods include ZDOCK (R. Chen et al. 2003; Pierce et al. 2014), HADDOCK (Dominguez et al. 2003), and ClusPro (Comeau et al. 2004). Despite their broad applicability, classical docking approaches are primarily limited by the need for accurate structures of the individual binding partners.

Recent advances in deep learning have substantially transformed protein complex prediction, leading to the development of AlphaFold-Multimer (Evans et al. 2022). Building upon the AlphaFold2 framework, AlphaFold-Multimer predicts protein complexes by incorporating inter-chain residue relationships into the alignment and pairwise representations, thereby enabling the model to learn inter-chain contact patterns directly (Evans et al. 2022; Rosenhouse-Dantsker 2026). However, prediction accuracy for complexes remains lower than for monomeric proteins, and the method is not designed for nucleic acid-containing or protein-ligand complexes (Evans et al. 2022). In addition, the size of complexes that can be modeled is limited by computational and memory constraints. In this thesis, AlphaFold2 refers to AlphaFold-Multimer unless stated otherwise.

ColabFold (Mirdita et al. 2022) was developed to improve the accessibility of AlphaFold2 and AlphaFold-Multimer by reducing the computational cost of MSA generation. It replaces the original JackHMMER and HHblits pipeline with MMseqs2 (Steinegger and Söding 2017), enabling faster structure prediction, especially for protein complexes.

AlphaFold3 (Abramson et al. 2024) extends the AlphaFold framework beyond protein-protein interactions to a wider range of biomolecular systems, including protein-DNA, protein RNA, protein-small molecule, and nucleic acid-only complexes, as well as proteins containing post-translational modifications. The model introduces a diffusion-based structure module that generates all atom coordinates and replaces the AlphaFold2 structure module (Abramson et al. 2024; Rosenhouse-Dantsker 2026). Moreover, the Evoformer module is replaced by a simplified architecture referred to as the Pairformer. AlphaFold3 is trained on a substantially broader set of structural data types and achieves high accuracy across multiple complex classes, including protein-nucleic acid and small-molecule interactions (Abramson et al. 2024).

Both AlphaFold2 and AlphaFold3 provide confidence metrics to assess prediction quality, including the predicted aligned error (PAE) and the predicted local distance difference test score (pLDDT) (Figure 1.8B). These metrics are described in detail in the subsequent sections.

Other successful end-to-end deep-learning protein structure prediction methods include RoseTTAFold (Baek et al. 2021), RoseTTAFold2 (Baek et al. 2023a),

RoseTTAFoldNA (Baek et al. 2023b) for protein-nucleic acid complexes, and the protein language model ESMFold (Z. Lin et al. 2023). As this thesis focuses on AlphaFold-based approaches, these methods are not described in detail.

Collectively, AlphaFold-based and other deep learning approaches have enabled large-scale computational modeling of biomolecular complexes across a wide range of interaction types (Varadi et al. 2022; L. Chen et al. 2024).

To make AlphaFold predictions broadly accessible, the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>) has complemented the PDB by providing predicted structural models at the proteome scale, thereby substantially extending the range of available structural information beyond experimentally determined structures (Varadi et al. 2022). However, the PDB remains the primary reference for experimentally validated structures and, therefore, for assessing the quality of computationally predicted complex structures.

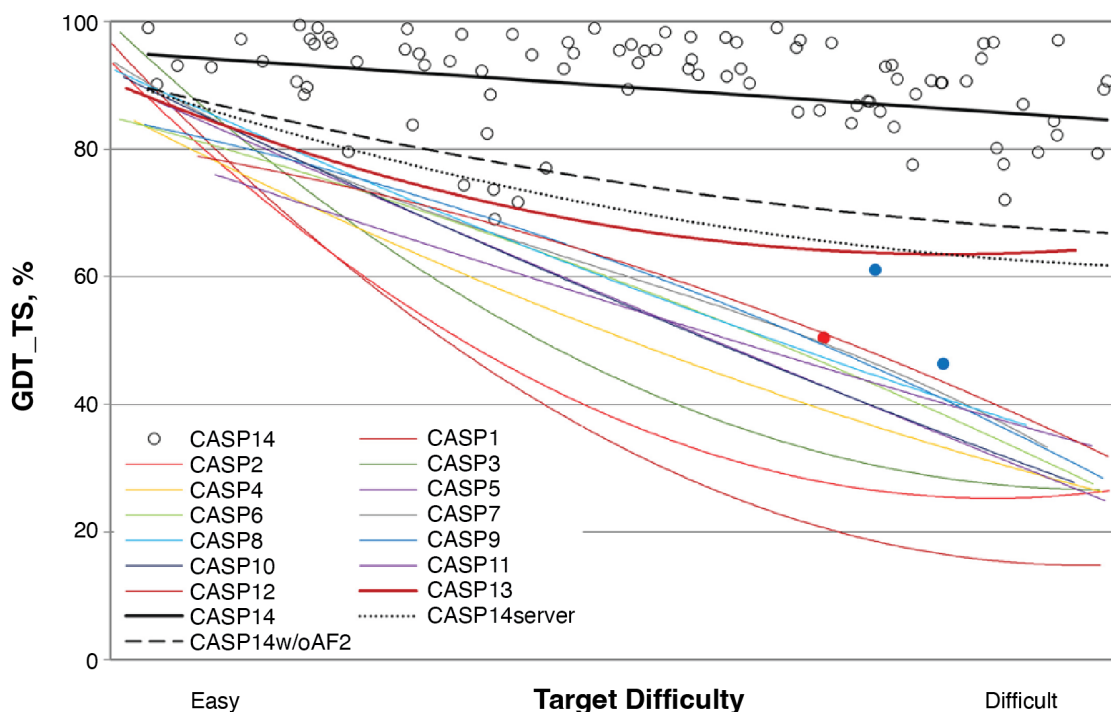


Figure 1.9: Backbone agreement with experiment across 14 CASP rounds. Trend lines show GDT_TS scores (see subsection 1.4.3) of the best models across CASP rounds 1 to 14, grouped by target difficulty. GDT_TS measures the agreement between predicted and experimental $C\alpha$ atom positions. Three targets with notably low agreement are highlighted. CASP14 results approach experimental accuracy, reflecting the impact of AlphaFold2. Adapted from Kryshchuk et al. (2021).

1.4 Analysis and evaluation of protein-protein interfaces

Characterizing protein-protein interfaces from experimental or predicted structures requires specialized computational methods for their identification, description, and evaluation. This section provides an overview of molecular visualization tools, summarizes approaches for interface identification and clustering, introduces metrics for assessing the accuracy of predicted complexes, and reviews scoring methods

for AlphaFold2 and AlphaFold3 predictions. It concludes by highlighting existing limitations that motivated the development of PICKCLUSTER and the benchmarking study presented in this thesis.

1.4.1 Molecular visualization of protein structures and complexes

Molecular visualization software is an essential component of structural biology workflows, enabling researchers to analyze three-dimensional structures, identify structural features, generate mechanistic hypotheses, and communicate findings (O'Donoghue et al. 2010; Verma et al. 2025). As structural datasets continue to expand in scale and complexity, including both experimentally determined structures and computationally predicted models, specialized visualization platforms have become increasingly important for structural analysis, interpretation, and reproducible research workflows.

RasMol (Sayle and Milner-White 1995) was among the first widely accessible molecular visualization programs and helped establish conventions for color-coded atom and secondary-structure representations. VMD (Humphrey et al. 1996) became widely adopted for visualizing and analyzing molecular dynamics simulations, while PyMOL (Schrödinger, LLC 2015) emerged as one of the most widely used tools in structural visualization. UCSF Chimera (Pettersen et al. 2004) introduced an extensible environment for interactive structural analysis, while UCSF ChimeraX (Goddard et al. 2018; Pettersen et al. 2021) was developed as a molecular visualization platform building upon Chimera to support integrated multiscale modeling and the interactive visualization, analysis, and interpretation of structural data. ChimeraX supports large-scale structure analysis, cryo-EM density visualization, and integrative modeling workflows. Several molecular visualization platforms, including PyMOL, VMD, Chimera, and ChimeraX, additionally enable the development of custom extensions and analysis plug-ins tailored to specific applications. In parallel, web-based visualization tools such as Mol* (Sehnel et al. 2021), used by the PDB and the AlphaFold Protein Structure Database, enable interactive visualization and analysis of structures directly in web browsers.

1.4.2 Tools for protein complex interface identification, analysis, and clustering

Various computational tools have been developed for the identification and characterization of protein-protein interfaces. These approaches differ in the features they compute and the level of analysis they provide.

PISA (Krissinel and Henrick 2007) is a widely used method for the analysis of interfaces in a crystal. It detects biological assemblies from PDB structures by assessing the thermodynamic stability of interfaces and calculating properties such as buried surface area (BSA), hydrogen bonds, and salt bridges. PISA is available as an interactive web server, PDBePISA, which provides precomputed results for all PDB entries and allows the analysis of user-submitted structures. InterfaceAnalyzer, part of the Rosetta modeling suite (Kaufmann et al. 2010; Bender et al. 2016), complements this by calculating interface binding energies, buried solvent accessible surface areas, and other properties for interfaces involving multiple chains. Similarly, PRODIGY (Xue et al. 2016) estimates binding affinities directly from structural features of protein-protein interfaces.

Several tools focus on the geometric and physicochemical characterization of protein-protein interfaces rather than thermodynamic properties. PSAIA (Protein Structure and Interaction Analyzer) (Mihel et al. 2008) computes properties such as solvent accessibility, atomic depth, and interaction types, including hydrophobic, polar, ionic, and van der Waals contacts. COCOMAPS (Vangone et al. 2011) and COCOMAPS 2.0 (Chawla et al. 2025) generate intermolecular contact maps for protein-protein, protein-DNA, and protein-RNA complexes based on user-defined distance thresholds. In addition, they report physicochemical properties of interacting residues, including multiple non-covalent interaction categories. Tools such as Arpeggio (H. C. Jubb et al. 2017) and PLIP (Schake et al. 2025) provide detailed annotation of interatomic interactions. PPI3D (Dapkūnas et al. 2016, 2024) enables comparison of interfaces across the PDB using Voronoi tessellation, which is a geometric partitioning method that defines spatial contacts based on nearest-neighbor relationships. This approach allows for structure-based searches for related interfaces and the identification of conserved contact patterns. VoroContacts (Olechnovič and Venclovas 2021) provides a Voronoi tessellation-based framework for the calculation of interatomic contacts at macromolecular interfaces, including contact areas and interaction annotations. PPIclust (Aung et al. 2008) clusters protein-protein interfaces extracted from three-dimensional structures, organizing them by similar interface patterns and enabling the identification of recurring interface geometries across the PDB.

1.4.3 Reference-based quality metrics

Assessing the accuracy of the predicted structure of a protein complex requires metrics that can quantify the agreement between a model and a reference, namely an experimental structure, at both global and local interface levels. Several measures have been established for this purpose, each capturing different aspects of structural similarity.

The CAPRI (Critical Assessment of Prediction of Interactions) criteria provide a widely used framework for assessing protein-protein docking and complex prediction (Méndez et al. 2003; Lensink et al. 2007; Lensink and Wodak 2013). Predicted models are assessed using a set of complementary metrics (Méndez et al. 2003; Lensink et al. 2007). The fraction of native contacts (F_{nat}) measures the proportion of residue contacts in the reference interface that are correctly reproduced in the model. In contrast, the fraction of non native contacts ($F_{\text{non-nat}}$) quantifies the proportion of contacts present in the model interface that are absent in the reference. The interface RMSD (I-RMSD) evaluates the root mean square deviation (RMSD) of backbone atoms of interface residues common to the model and reference after superposition of the interface region. The ligand RMSD (L-RMSD) measures the backbone RMSD of the smaller chain after superimposing the larger chain of the complex. In addition, steric clashes are assessed as contacts between non-hydrogen atoms at distances below 3 Å (Méndez et al. 2003).

DockQ and DockQ v2 (Basu and Wallner 2016; Mirabello and Wallner 2024) provide a single continuous quality score for protein-protein docking and prediction models by combining F_{nat} , I-RMSD, and L-RMSD into a single metric ranging from 0 for incorrect models to 1 for perfect predictions:

$$\text{DockQ}(F_{\text{nat}}, \text{LRMS}, \text{iRMS}) = \frac{F_{\text{nat}} + \text{RMS}_{\text{scaled}}(\text{LRMS}, d_1) + \text{RMS}_{\text{scaled}}(\text{iRMS}, d_2)}{3} \quad (1.2)$$

The function $\text{RMS}_{\text{scaled}}$ applies a non-linear scaling to the RMSD values using scaling parameters optimized to $d_1 = 8.5 \text{ \AA}$ and $d_2 = 1.5 \text{ \AA}$ (Basu and Wallner 2016), ensuring comparable contributions of all terms. Established DockQ classification thresholds are commonly used to assess the quality of protein-protein docking models and protein complex predictions (Table 1.1) (Basu and Wallner 2016). The metric serves as the reference measure in this thesis.

Table 1.1: DockQ classification for protein complex models based on CAPRI criteria (Basu and Wallner 2016).

DockQ range	Model classification
$\text{DockQ} < 0.23$	Incorrect
$0.23 \leq \text{DockQ} < 0.49$	Acceptable quality
$0.49 \leq \text{DockQ} < 0.80$	Medium quality
$\text{DockQ} \geq 0.80$	High quality

In addition to CAPRI metrics and DockQ, several structural similarity metrics are commonly used, particularly in CASP evaluations (Kryshtafovych et al. 2019; Studer et al. 2023). The TM score (Template Modeling score) (Y. Zhang and Skolnick 2004) is a length-normalized global measure of structural similarity that ranges from 0 to 1, with values above 0.5 typically indicating the same overall fold. The local distance difference test (IDDT) (Mariani et al. 2013) is a superposition-free per-residue metric that evaluates differences in local interatomic distance patterns between a model and a reference structure. The GDT_TS (Global Distance Test Total Score) (Zemla 2003; W. Li et al. 2016) measures the fraction of residues of the target and the reference structure that can be superimposed within predefined distance thresholds and is one of the primary ranking metrics in CASP (Kryshtafovych et al. 2019). Conveniently, platforms such as OpenStructure (Biasini et al. 2013) support the automated calculation of these metrics, including IDDT and interface-specific scores.

1.4.4 Scoring metrics for AI-generated protein complex predictions

The metrics described in the preceding subsections are reference-dependent, as they evaluate the quality of predicted models by comparison with experimentally determined structures. With the rapid development of AI-based protein complex prediction, however, the majority of generated models lack an experimental reference. This increases both the difficulty and the importance of reliable model quality estimation.

AlphaFold2 and AlphaFold3 address this limitation by providing confidence metrics at multiple levels, including residue, atom, model, and pairwise representations. These scores serve as primary indicators of model reliability, as they can be computed for any AlphaFold-based prediction without requiring a reference structure.

The predicted local distance difference test (pLDDT) (Jumper et al. 2021) is a confidence score reported per residue in AlphaFold2 and per atom in AlphaFold3, ranging from 0 to 100. It estimates the expected IDDT $C\alpha$ score (Mariani et al. 2013) relative to a hypothetical experimental structure (Figure 1.8B). Values above 90 indicate high confidence and are typically associated with well-ordered and accurately positioned residues. Scores between 70 and 90 reflect generally reliable backbone geometry. Values between 50 and 70 suggest lower confidence, often corresponding to

flexible or weakly constrained regions, while values below 50 are commonly associated with intrinsically disordered regions or unreliable predictions (Jumper et al. 2021; Gomez and Kovalevskiy 2025b). For the evaluation of protein-protein interfaces, the mean pLDDT over interface residues, referred to as interface pLDDT (ipLDDT), has been used to assess the confidence of predicted protein-protein interfaces (Bryant et al. 2022).

The predicted aligned error (PAE) (Evans et al. 2022) provides pairwise confidence estimates represented as an $N \times N$ matrix, where N denotes the number of residues (Figure 1.8B). Each entry $\text{PAE}(i, j)$ corresponds to the expected positional error of residue j , defined as the distance between its $\text{C}\alpha$ atom in the predicted structure and the corresponding atom in the experimental structure, after superposition of the backbone atoms of residue i (Jumper et al. 2021; Abramson et al. 2024). During training, when the experimental structure is available, the network learns to predict a probability distribution over this aligned error distance. During prediction, when no experimental structure is available, the PAE value is then obtained as the expected value of this predicted distribution and is capped at 31.75 Å. Low PAE values indicate low error, and therefore high confidence in the relative residue positioning of residue j given residue i . For multimeric complexes, the PAE matrix contains both intra-chain and inter-chain entries. Intra-chain entries, where residues i and j belong to the same polypeptide chain, reflect confidence in the internal fold of each subunit. In contrast, inter-chain entries, where i and j belong to different chains, reflect confidence in the relative positioning of those chains within the complex and are therefore particularly informative for assessing predicted protein-protein interfaces (Evans et al. 2022).

PAE is used to derive the predicted TM score (pTM) (Jumper et al. 2021), a global model accuracy estimate ranging from 0 to 1, and the interface pTM score (ipTM), which focuses on inter-chain regions. For ipTM, AlphaFold provides only limited guidance, identifying values between 0.6 and 0.8 as a gray zone in which the model cannot be reliably classified as correct or incorrect (Gomez and Kovalevskiy 2025a). A combined *model confidence* score is defined as $0.8 \cdot \text{ipTM} + 0.2 \cdot \text{pTM}$, reflecting the emphasis on interface accuracy in complex prediction (Evans et al. 2022; Abramson et al. 2024).

In addition to the intrinsic AlphaFold confidence scores, external metrics have been developed to assess interface quality in the absence of reference structures. VoroIF-GNN (Olechnovič and Venclovas 2023) is a recent method that uses Voronoi tessellation and graph representations to estimate interface accuracy at the contact level. To compute the interface-specific score, predicted residue level accuracy values are combined into an average VoroIF-GNN residue pCAD score. This score is subsequently adjusted to account for structural completeness based on the proportion of atoms present in the predicted model during training. This metric ranked among the top-performing methods for interface quality estimation in the CASP15 Estimation of Model Accuracy (EMA) category (Studer et al. 2023). Other high-performing methods included ModFOLDdock and ModFOLDdockR (Edmunds et al. 2023), as well as the VoroMQA metric (Olechnovič and Venclovas 2019). In CASP16, ModFOLDdock2 and ModFOLDdock2R (McGuffin et al. 2025) achieved top performance. Notably, the AlphaFold3 web server also showed strong performance in CASP16 (Fadini et al. 2026). The predicted DockQ score (pDockQ) (Bryant et al. 2022) provides an alternative approach by combining interface contacts and the mean confidence of the interacting residues into a sigmoid function calibrated against DockQ (Basu and Wallner 2016). Its successor, pDockQ2, was introduced to improve evaluation of multimeric assemblies (Zhu et al. 2023).

1.4.5 Benchmarking initiatives in structural model evaluation

Community benchmarking initiatives have been central to developing and validating both reference-dependent metrics and quality estimation methods, and provide the standardized datasets and protocols used in this thesis.

As part of the CASP competition described in subsection 1.3.1, the Estimation of Model Accuracy (EMA) category, introduced at CASP7, established the blind assessment of model quality as a distinct challenge separate from structure prediction (Cozzetto et al. 2007). EMA methods have been evaluated in each subsequent CASP round, with substantial efforts to integrate structural, evolutionary, and statistical features into accurate quality predictions (Studer et al. 2023; Fadini et al. 2026). CAPRI (Critical Assessment of PRedicted Interactions) provides a complementary framework focused on benchmarking protein-protein docking and complex structure prediction (Janin 2005; Lensink et al. 2007; Lensink and Wodak 2013).

In parallel, continuous, automated approaches for evaluating prediction methods were developed. These included the EVA server (Eyrich et al. 2001), which assessed publicly available prediction servers against newly released PDB structures on a weekly basis. CAMEO (Continuous Automated Model Evaluation) (Robin et al. 2026) subsequently expanded this concept by introducing a continuous, blind evaluation pipeline based on newly determined experimental structures prior to their release.

RNA Puzzles (Cruz et al. 2012; Bu et al. 2025) have applied the same blind prediction paradigm to RNA structure prediction, highlighting both progress and remaining challenges in nucleic acid modeling. More recently, nucleic acid-containing targets and protein-nucleic acid complexes have been incorporated into CASP, further extending community-wide benchmarking to these systems (Kretsch et al. 2023, 2026). The community study by Schweke et al. (Schweke et al. 2023) evaluated methods for distinguishing physiological from non-physiological interfaces in PDB structures, emphasizing the challenge of differentiating biologically relevant interactions from crystal packing artifacts.

These initiatives have directly shaped the metrics and benchmarking practices used in this thesis.

1.4.6 Motivation for PICKLUSTER

As outlined in the preceding sections, protein-protein interfaces are structurally and chemically heterogeneous. As discussed in subsection 1.2.2, hotspot residues tend to cluster within spatially separated sub-regions rather than being uniformly distributed across the interface (Clackson and Wells 1995; Keskin et al. 2005). In addition, large interfaces often consist of multiple distinct contact patches arising from interactions of different structural domains (Keskin et al. 2005). Consequently, the biological and therapeutic relevance of an interface cannot be fully captured by treating it as a single entity. Therefore, the decomposition into spatially discrete sub-interfaces provides an approach to identify regions most critical for binding, enriched in hotspot residues, or relevant for structure-guided drug design (Blundell et al. 2000; H. Jubb et al. 2015). However, existing interface analysis tools described in subsection 1.4.2 lack the capability to perform spatial decomposition into distinct sub-regions.

The rapid growth of predicted protein complexes generated by AlphaFold2 and AlphaFold3 (Evans et al. 2022; Abramson et al. 2024) has fundamentally changed

structural analysis. The majority of newly available complex structures are now predicted rather than experimentally determined and are accompanied by confidence metrics such as pLDDT, PAE, and ipTM, which quantify model quality in local structure and inter-chain arrangement. As outlined in subsection 1.4.4, these scores are essential for downstream analyses and interpretation. High-confidence interfaces differ fundamentally from low-confidence ones, and clusters of low-confidence residues may indicate structural uncertainty rather than defined binding regions. Therefore, analyses that do not account for prediction confidence risk treating all residues equally, regardless of their structural reliability. However, classical interface analysis tools such as PISA and InterfaceAnalyzer do not incorporate prediction-based confidence measures into their workflows.

Moreover, most interface analysis tools are implemented as command-line programs or web servers, often requiring separate processing and interpretation of results outside the structural context. While the AlphaFold Database and ChimeraX allow visualization of predicted structures and identification of inter-chain contacts, they do not support identification of sub-interface regions based on spatial separation or the integration of residue-level confidence metrics with identified interfaces.

Together, these three gaps, the lack of a tool for spatial sub-interface decomposition, the absence of confidence score integration, and the lack of an interactive solution in a visualization environment, motivated the development of PICKLUSTER (Genz et al. 2023). PICKLUSTER is an open-source ChimeraX plug-in and command-line tool that performs distance-based spatial clustering of interface residues into sub-interfaces. It operates on both experimentally determined and AlphaFold-predicted structures and enables direct visualization of residue-level confidence scores within identified contact regions, enabling confidence-aware interface analysis within an interactive visualization environment.

1.4.7 Motivation for the benchmarking of scoring metrics and C2Qscore

The confidence and interface quality metrics described in subsection 1.4.4 are widely used in structural biology. However, a fundamental question remains unresolved concerning which metric most accurately reflects the true structural quality of predicted protein-protein interfaces and at what score threshold such predictions can be considered reliable. Without systematic benchmarking against a consistent ground truth, researchers lack a basis for selecting among available metrics when evaluating predicted complexes. This challenge is reinforced by the absence of scoring thresholds, as most metrics provide continuous values without binary classification cutoffs. The derivation of robust thresholds, therefore, requires a systematically curated benchmark dataset with experimentally determined reference structures and a consistent ground-truth measure, such as DockQ.

The available benchmarking evidence is limited, as most metrics have been developed and validated independently on different datasets and with distinct reference measures. The study by Yin et al. (2022) evaluated pLDDT, pTM, and an interface-specific PAE against I-RMSD for 152 heterodimeric complexes and compared AlphaFold2 with the docking program ZDOCK, providing insights into the relationship between confidence scores and inter-chain accuracy. However, this work predates the release of AlphaFold3.

Although binary classification thresholds were proposed, they were derived from a relatively small dataset and did not include externally developed scoring methods. In addition, the study employed an interface definition based on a 4 Å heavy-atom distance cutoff, which differs from the criteria used by the AlphaFold server and may exclude functionally relevant residues. More recent methods, such as VoroIF-GNN (VoroIF) and pDockQ2, have been validated on separate datasets (Olechnovič and Venclovas 2023; Zhu et al. 2023).

These gaps motivated the systematic benchmarking study presented in this thesis (Genz et al. 2025a). A curated benchmark comprising 223 high-resolution heterodimeric protein complexes from the PDB is used to evaluate seven scoring metrics, including ipLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF, across the three prediction methods ColabFold with templates (CF-T), ColabFold without templates (CF-F), and AlphaFold3 (AF3). DockQ is employed as the reference measure as it is widely adopted in the structural biology community, and provides both a continuous quality score and well-defined classification thresholds (Basu and Wallner 2016). Receiver operating characteristic (ROC) curve analysis is used to derive binary classification thresholds for each metric and scoring method. Based on these results, C2Qscore is introduced as a combined quality score integrating five of the individual metrics (ipLDDT, iPAE, pTM, ipTM, and VoroIF) into a single metric to assess interface accuracy. It is validated on two independent datasets derived from X-ray crystallography and cryo-EM. Together, the identified best-performing metrics, the derived thresholds, and C2Qscore provide a practical framework for assessing the reliability of predicted protein complexes in the absence of experimental reference structures.

1.5 Beyond proteins: Expanding to nucleic acid-containing complexes

The preceding sections have focused on protein-protein interfaces, their structural principles, and the computational tools and metrics used for their analysis and quality assessment. However, as described in subsection 1.1.1, proteins also form functionally important interactions with nucleic acids. These assemblies, as well as interactions among nucleic acids themselves, are essential for a wide range of biological processes, including transcription, splicing, translation, and genome replication, and represent a significant yet comparatively less well-characterized class of therapeutic targets. This section introduces the structural and biological features of nucleic acid-containing assemblies, outlines the experimental and computational challenges specific to these complex classes, and reviews the available tools for their analysis. It concludes by highlighting the limitations that motivated the development of the tool ProNA3D, which is presented in this thesis.

1.5.1 Biological roles and structural features of nucleic acid-containing complexes

Nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are the primary carriers of genetic information and key mediators of its expression (Branden and Tooze 1999; Liljas et al. 2009; Voet et al. 2016). The directional flow of information from DNA

via RNA to protein constitutes the central dogma of molecular biology (F. Crick 1970). Both DNA and RNA are linear polymers composed of nucleotides, each consisting of a pentose sugar (ribose in RNA or 2'-deoxyribose in DNA), a phosphate group attached to the 5' carbon, and a nitrogenous base linked to the 1' carbon (Figure 1.10) (Voet et al. 2016). Nucleotides are connected through phosphodiester bonds linking the 3' and 5' positions of adjacent sugars, forming the nucleic acid backbone. DNA contains the bases adenine (A), guanine (G), cytosine (C), and thymine (T), whereas RNA contains uracil (U) in place of thymine (Liljas et al. 2009). Watson Crick base pairing, in which adenine pairs with thymine or uracil and guanine pairs with cytosine via specific hydrogen bonding patterns, provides the molecular basis for the double helical structure of DNA and contributes to the formation of defined secondary structures in RNA (Figure 1.10) (Watson and F. H. Crick 1953; Branden and Tooze 1999; Liljas et al. 2009). While DNA predominantly adopts a stable double helical conformation, RNA is typically single-stranded and exhibits a wide range of secondary and tertiary structures, including stem loops, bulges, and pseudoknots, which are closely linked to its diverse functional roles (Liljas et al. 2009). Similar to proteins, nucleic acid structure can be described at the primary, secondary, tertiary, and quaternary level (Liljas et al. 2009). Secondary structure is primarily determined by base pairing interactions, including both canonical (Watson-Crick) and non-canonical pairs, with the latter being particularly prevalent in RNA, contributing to its structural diversity (Liljas et al. 2009; Voet et al. 2016). Base stacking interactions, in which adjacent bases align in a nearly parallel orientation and interact predominantly via π - π and van der Waals contacts, provide a major contribution to structural stability. Tertiary structure arises from the spatial arrangement of secondary structural elements and is stabilized by hydrogen bonding, van der Waals interactions, and electrostatic forces.

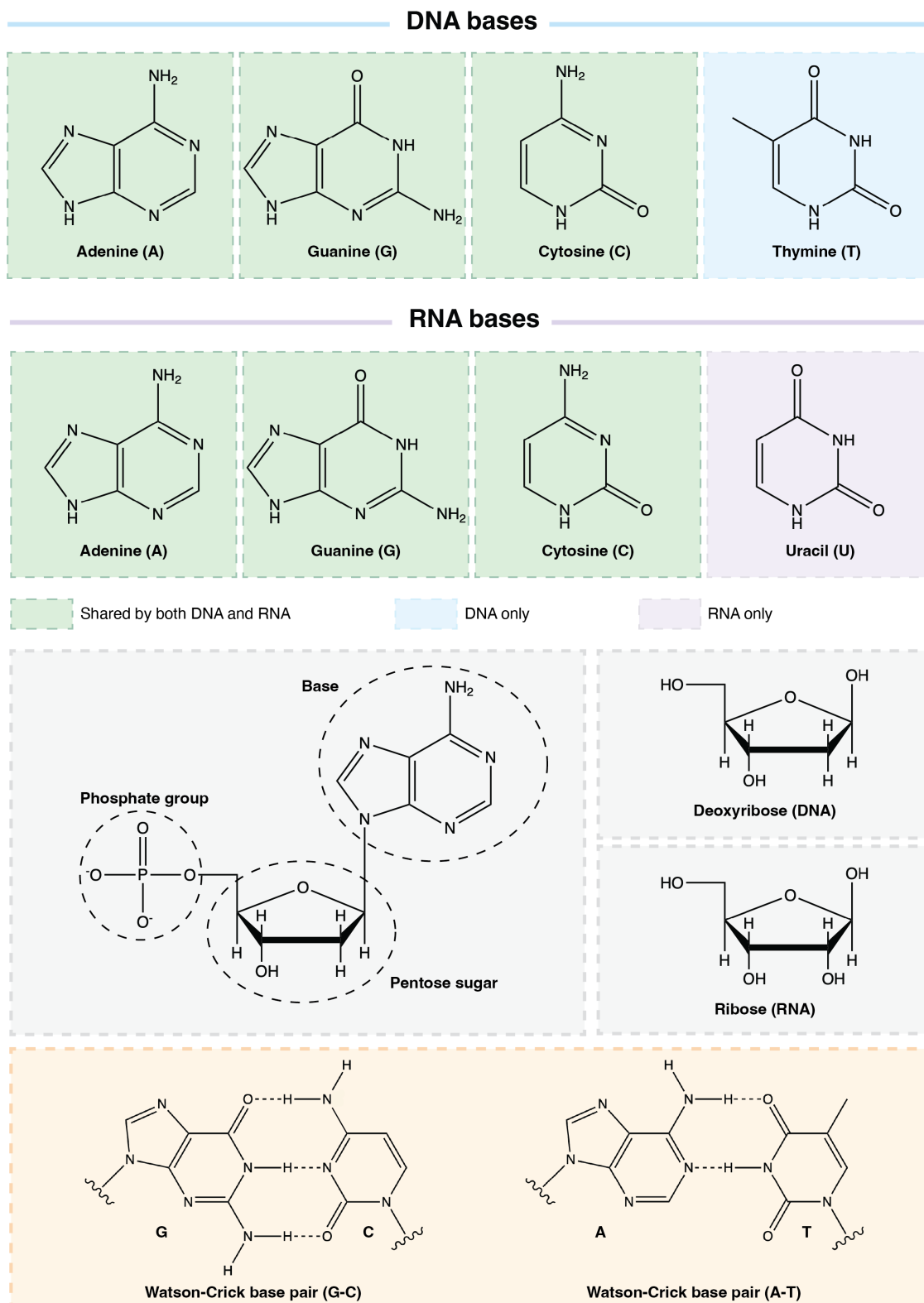


Figure 1.10: Overview of nucleic acid building blocks, nucleotide composition, and canonical Watson-Crick base pairs.

Proteins interact with both DNA and RNA and can contribute to the stabilization of the tertiary structure of nucleic acids (Liljas et al. 2009). These interactions are essential for gene expression and genome maintenance, including processes such as DNA replication, repair, transcription, and RNA metabolism (Doudna and Richmond 2001; Liljas et al. 2009; Hudson and Ortlund 2014). Protein-DNA interactions are central to DNA replication, where polymerases and clamp proteins engage the DNA template (Liljas et al. 2009). They are also crucial for transcription, as transcription factors and RNA polymerases recognize specific promoter elements, and for DNA repair, where repair complexes must locate and process damaged bases (N. Tuteja and R. Tuteja 2001; Alberts 2003; Ptashne 2005; Liljas et al. 2009). Protein-RNA interactions are equally important. The ribosome is a protein-RNA assembly in which ribosomal proteins stabilize the catalytic RNA scaffold responsible for protein synthesis (Liljas et al. 2009). Splicing is mediated by the spliceosome, a complex of small nuclear RNAs and associated proteins (Hertel and Graveley 2005). In addition, post-transcriptional regulation relies on RNA-binding proteins that regulate transcript stability, localization, and translation (Halbeisen et al. 2008; Khoroshkin et al. 2024).

The structural principles of nucleic acid recognition differ from those of protein-protein interactions. The regular phosphate backbone, the pattern of hydrogen bond donors and acceptors in DNA grooves, and the structural diversity of RNA provide specific binding features for proteins (Jones et al. 1999, 2001). Protein-DNA interfaces are typically characterized by a high number of direct and water-mediated hydrogen bonds, combined with hydrophobic contacts on the protein side (Jones et al. 1999; Rohs et al. 2010). In contrast, protein-RNA recognition must account for the greater conformational flexibility and structural diversity of RNA, which presents complex three-dimensional architectures for binding. In such interfaces, van der Waals interactions are often more prominent than direct hydrogen bonding (Draper 1995; Jones et al. 2001). Electrostatic interactions between positively charged protein residues and the negatively charged phosphate backbone generally drive nucleic acid binding, while base-specific contacts and shape complementarity provide binding specificity (Jones et al. 1999; Anashkina 2023). Despite these differences, protein-nucleic acid binding is often governed by hotspot residues contributing the majority of the binding energy, as observed in protein-protein interactions (Pan et al. 2018, 2020; Anashkina 2023).

1.5.2 Challenges in structural determination and prediction of nucleic acid-containing complexes

The experimental determination of nucleic acid-containing complex structures presents challenges that are distinct from those encountered for protein-only assemblies. Although X-ray crystallography and cryo-EM remain the primary methods for structure determination, both are challenged by the properties of nucleic acids, particularly RNA. RNA exhibits high conformational flexibility and adopts ensembles of structures in solution rather than a single well-defined structure, which hinders the formation of well-ordered crystals (Ke and Doudna 2004; Kappel et al. 2020). In addition, the highly negatively charged phosphate backbone poses additional challenges for crystal packing and increases the likelihood of crystallographic artifacts (Ferré-D'Amaré et al. 1998; Ke and Doudna 2004).

In cryo-EM, the conformational heterogeneity and flexibility of RNA-containing assemblies pose substantial challenges for particle classification and reconstruction, often limiting the achievable resolution in flexible regions (Bonilla and Kieft 2022). Consequently, nucleic acid-containing complexes are underrepresented in the PDB relative to their biological importance. As shown in Figure 1.6, they account for less than 9% of all entries, compared to more than 91% for protein-only structures (Ma et al. 2022).

Computational prediction methods have begun to address the limited structural coverage, but the accuracy achieved for nucleic acid-containing assemblies remains substantially lower than that for protein-only complexes (Kretsch et al. 2026; Kryshtafovych et al. 2026). RoseTTAFoldNA (Baek et al. 2023b) expanded deep learning approaches to protein-nucleic acid complexes by modifying the RoseTTAFold architecture to incorporate RNA and DNA. More recently, AlphaFold3 (Abramson et al. 2024) was developed, which represents the most comprehensive approach to date. It employs a diffusion-based architecture across diverse biomolecular interaction types and achieves improved accuracy for protein-nucleic acid complexes compared to earlier methods (Peng et al. 2025). Despite these advances, prediction performance for nucleic acid-containing assemblies remains significantly lower than for protein-only predictions (Kryshtafovych et al. 2026). This limitation was highlighted at CASP16, in which the assessment of nucleic acid structure prediction showed that the best performing models were generated by human expert groups rather than automated methods (Kretsch et al. 2026). This reflects both the complexity of nucleic acid conformations and the limited availability of high-quality training data. As a result, predicted models require careful interpretation, and the integration of confidence metrics is particularly important for a reliable analysis.

1.5.3 Computational tools for the analysis of nucleic acid-containing complexes

Various computational tools have been developed to analyze nucleic acid-containing complexes, including methods for secondary structure annotation, interaction characterization, and interface visualization (Figure 1.11).

Several tools support RNA secondary structure annotation and visualization. ViennaRNA (Lorenz et al. 2011) provides a widely used suite of algorithms for RNA secondary structure analysis, including the prediction of minimum free energy structures and the calculation of base pair probabilities from sequence information. RNAView (H. Yang et al. 2003) analyzes three-dimensional structures to identify base pairing interactions in RNA and DNA, enabling systematic annotation of canonical and non-canonical pairs and their representation in two-dimensional diagrams. Similarly, the web server RNAscape (Mitra et al. 2024) generates two-dimensional topology diagrams that map nucleotide positions from three-dimensional structures onto their secondary structures, facilitating the interpretation of complex RNA and DNA folds. DSSR (Dissecting the Spatial Structure of RNA) (Lu et al. 2015) provides detailed structural annotations, including base pair classification and the identification of structural motifs such as junctions and pseudoknots, and serves as a basis for several downstream tools, such as RNAscape. Visualization programs such as VARNA (Darty et al. 2009) and forna (Kerpedjiev et al. 2015) offer flexible two-dimensional representations of RNA secondary structure independently of a three-dimensional model.

Dedicated databases and analysis pipelines have been developed for the systematic description and analysis of nucleic acid-containing complex interfaces. DNAproDB (Mitra et al. 2025a) and RNAproDB (Mitra et al. 2025b) provide curated interface features for protein-DNA and protein-RNA complexes from the PDB and support user-submitted structures. Both platforms integrate nucleotide annotations based on DSSR, secondary-structure topology diagrams generated by RNAscape, and interactive three-dimensional visualizations. More general interface analysis tools, as described in subsection 1.4.2, including PLIP (Adasme et al. 2021), PPI3D (Dapkūnas et al. 2024), COCOMAPS (Vangone et al. 2011; Chawla et al. 2025), and VoroContacts (Olechnovič and Venclovas 2021), were extended to also support the characterization of contacts in nucleic acid-containing interfaces. The specialized tool ProDFace (Pal et al. 2022) provides detailed physicochemical and geometric descriptions of protein-DNA interfaces, including contact maps and residue-level interaction features.

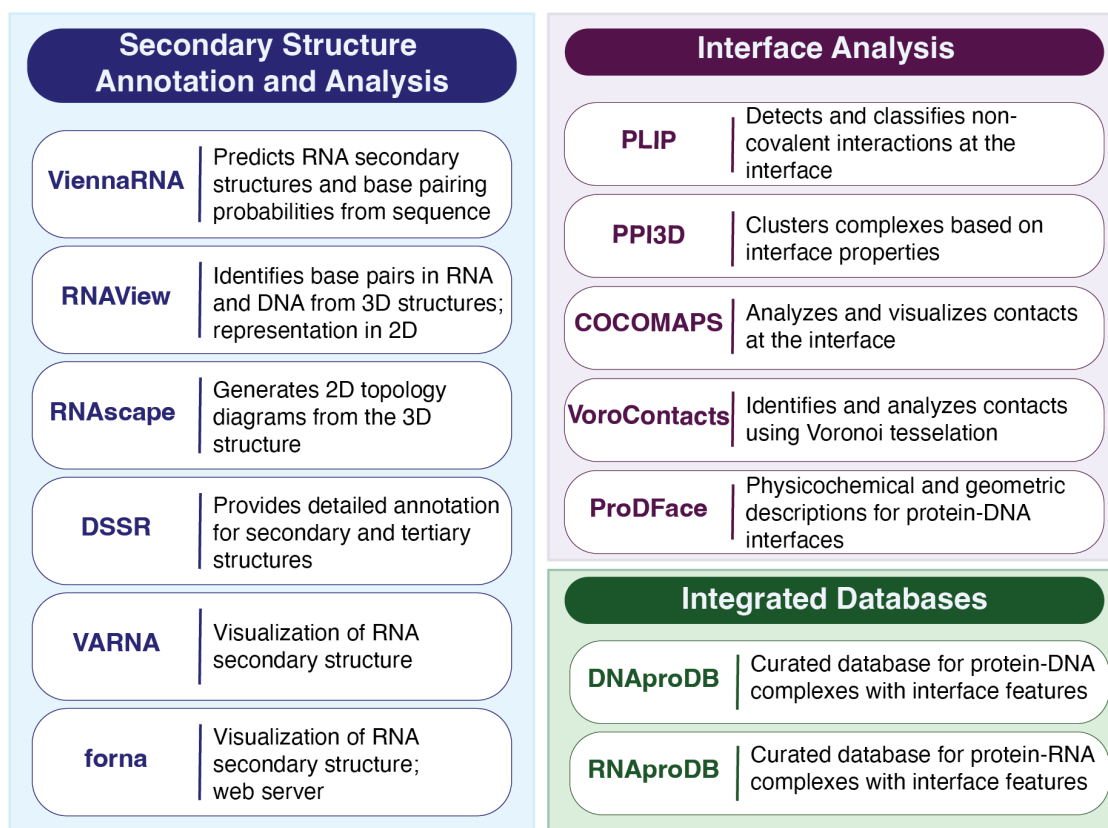


Figure 1.11: Overview of computational tools for analyzing the structure of nucleic acid-containing complexes. The tools are grouped based on their main application.

1.5.4 Motivation for ProNA3D

Despite the range of tools described above, several key limitations remain. Most existing platforms have been developed for experimentally determined structures and do not support the integration of confidence metrics derived from structure predictions. This represents a critical limitation, given the substantially lower prediction accuracy for nucleic acid-containing complexes described in subsection 1.5.2.

Furthermore, many available tools have been developed independently for proteins, RNA, or DNA, resulting in fragmented workflows when analyzing mixed biomolecular assemblies (Dapkūnas and Venclovas 2025). In addition, many platforms are either web-based with fixed functionality or command-line tools without integrated visualization. Finally, none of the available tools provides an integrated framework that combines two-dimensional secondary structure topology representations with three-dimensional sub-interface identification and graph-based visualization of interface contact networks. Such integration would enable identification of hub nucleotides or residues analogous to hotspots in protein-protein interfaces. These limitations highlight the need for a unified framework within an interactive graphical environment that supports both experimental and computationally predicted structures across macromolecular complex types.

The limitations observed for nucleic acid-containing complexes largely mirror those that motivated the development of PICKLUSTER for protein-protein interfaces. ProNA3D was developed to address these challenges by extending the interface detection and spatial sub-interface clustering algorithm of PICKLUSTER to protein-nucleic acid and nucleic acid-only assemblies. It further integrates secondary structure topology visualization, two-dimensional interface contact graphs, and AlphaFold3 confidence metrics directly into the analysis workflow. ProNA3D supports cryo-EM density map zoning for the analysis of flexible interface regions, which are particularly prevalent in RNA-containing complexes. Implemented as both a ChimeraX plug-in and a command-line tool, ProNA3D provides an integrated and confidence-aware framework for the analysis of nucleic acid-containing interfaces.

Taken together, the background presented in this chapter establishes the conceptual and methodological foundation for the three aims of this thesis. The structural principles of protein-protein and nucleic acid-containing complexes and their interfaces, the capabilities and limitations of experimental structure determination, the advances and remaining challenges of AI-based structure prediction, and the fragmented state of existing interface analysis and quality assessment tools highlight a common need. Interactive, confidence-aware, and unified computational frameworks are required for the reliable characterization of interfaces across the full spectrum of macromolecular interaction types. The methods, results, and discussion presented in the following chapters describe how PICKLUSTER, the scoring metrics benchmarking study, C2Qscore, and ProNA3D collectively address these challenges.

2

Aim

The overall aim of this thesis is to develop and apply computational approaches for the systematic characterization and quality assessment of macromolecular interfaces across protein-protein, protein-RNA, protein-DNA, and nucleic acid-only complexes, while highlighting their utility in structural biology. This work is motivated by the rapid expansion of AI-generated structural models, which have created an urgent need for analytical frameworks capable of reliably assessing and interpreting predicted interfaces. Here, three interconnected aims were defined to address this need.

Aim 1: Development of PICKLUSTER for the analysis of protein-protein interfaces.

Existing interface analysis tools lack methods for spatial sub-interface decomposition, do not incorporate prediction-based confidence metrics, and are not integrated into interactive visualization environments such as ChimeraX (Pettersen et al. 2021) (see subsection 1.4.6). To address these limitations, this work focuses on the development of PICKLUSTER as a ChimeraX plug-in and command-line tool for the identification and clustering of protein-protein interface contacts.

Specifically, the central aim is to demonstrate how spatial decomposition of protein-protein interfaces can support both the detailed investigation of individual complexes to guide structural analyses and the large-scale characterization of interface properties across the PDB. Particular emphasis is placed on the identification of geometrically distinct interface regions that may be relevant for structure-guided drug discovery applications. In addition, this work aims to improve the interpretability of AlphaFold-based complex predictions through the integration and visualization of confidence scores, including interface-specific metrics for the evaluation of individual sub-interfaces.

PICKLUSTER has been published (Genz et al. 2023) and is described in detail in this thesis.

Aim 2: Systematic benchmarking of scoring metrics for AI-generated protein complex predictions and development of C2Qscore.

The absence of systematic benchmarking against a consistent reference score and the lack of empirically derived classification thresholds represent key limitations of existing interface quality metrics for protein complex predictions (see subsection 1.4.7). This

work addresses these gaps by performing a comprehensive benchmarking of commonly used confidence scores, including pLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF, across multiple prediction methods, including ColabFold with templates, ColabFold without templates, and AlphaFold3.

The primary aim is to investigate how these metrics perform across the different prediction methods and to determine which approaches provide the most reliable assessment of complex quality. A further objective is to identify structural characteristics of protein complexes that may be associated with successful modeling outcomes, thereby providing guidance for target selection. In addition, this work addresses whether the combination of scoring metrics yields a more robust approach to complex quality assessment than any individual score alone. Since most confidence metrics provide continuous values without established binary classification cutoffs, an additional objective is to derive prediction method- and metric-specific thresholds for distinguishing correct from incorrect complex predictions. Finally, this work aims to evaluate the applicability of the examined metrics and thresholds across different structural contexts, including heterodimeric X-ray crystal structures and dimers derived from larger cryo-EM assemblies.

The results of this benchmarking study have been published (Genz et al. 2025a) and are presented in detail in this thesis.

Aim 3: Development of ProNA3D for the analysis of nucleic acid-containing interfaces.

Existing analysis tools for nucleic acid-containing complexes are fragmented and lack integration of confidence-aware sub-interface analysis, secondary structure topology visualization, and interactive graphical access across diverse macromolecular interaction types (see subsection 1.5.4). To address these limitations, this work focuses on the development of ProNA3D as a ChimeraX plug-in and command-line tool that extends the sub-interface clustering framework of PICKLUSTER to protein-nucleic acid and nucleic acid-only interactions.

The central aim is to investigate whether spatial sub-interface decomposition can provide similarly informative insights for nucleic acid-containing complexes and thereby improve the understanding of their interface organization. In addition, this work seeks to demonstrate how the analysis of secondary-structure topology plots and interface connectivity networks can facilitate the identification of interaction hubs that may be relevant to complex stability and function. Further objectives include assessing the utility of ProNA3D's cryo-EM density zoning feature for the analysis of flexible regions in cryo-EM-derived complexes and improving the interpretation of AlphaFold3-based complex predictions through the integration and visualization of interface-specific confidence metrics. At a larger scale, this work aims to characterize structural properties across diverse nucleic-acid-containing interface classes in the PDB, including sub-interface size distributions, normalized B-factors, and connectivity patterns. The final objective is to investigate whether nucleic acid-containing interfaces in viral complexes exhibit distinct structural characteristics.

ProNA3D has been published as a preprint (Genz and Topf 2026) and is detailed in this thesis.

3

Results

Contents

3.1	PICKLUSTER: A Protein-Interface Clustering and Analysis Plug-in for UCSF ChimeraX	35
3.2	Benchmarking PICKLUSTER on Large-Scale Protein Complex Data	37
3.3	Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions	39
3.3.1	Benchmarking the prediction accuracy of ColabFold with templates, ColabFold without templates, and AlphaFold3	39
3.3.2	Influence of the heavy atom distance parameter for interface calculation	43
3.3.3	Distributions and correlation of individual prediction-based scores	43
3.3.4	Correlation of the individual prediction-based scores with DockQ	44
3.3.5	Determinants of high- and low-quality models	46
3.3.6	Combined scoring and testing	49
3.3.7	Evaluating the scores on dimers from larger assemblies	50
3.4	ProNA3D: Distance-Based Analysis of Nucleic Acid and Protein Interfaces	52
3.4.1	X-ray structure of a protein-RNA complex: <i>Acanthamoeba polyphaga</i> mimivirus RNA guided nuclease-RNA	54
3.4.2	X-ray structure of an RNA-only complex: hammerhead ribozyme	56
3.4.3	Predicted protein-RNA complex: human Fab antibody-HIV-1 RNA	58
3.4.4	Predicted protein-DNA complex: Cre recombinase from <i>Punavirus P1</i> bound to its target DNA	60
3.4.5	Cryo-EM application of ProNA3D	62
3.5	Systematic analysis of protein-nucleic acid and nucleic acid-only interfaces using ProNA3D	64
3.5.1	Interface size distributions across nucleic acid-containing complexes	64

3.5.2	Normalized B-factor analysis of interface regions	68
3.5.3	Maximum connectivity distributions for different interface classes	69
3.5.4	Individual residue- and nucleotide-level connectivity for different interface classes	74
3.5.5	Occurrence distributions of residue- and nucleotide-types for different interface classes	76
3.5.6	Nucleic acid-containing interfaces in viral complexes	80

The results presented in this chapter are organized around the three interconnected aims. First, PICKLUSTER is introduced as a ChimeraX plug-in and command-line tool for the identification, subdivision, and analysis of protein-protein interfaces. The tool is applied on a large benchmark dataset of protein complexes from the PDB to structurally characterize protein-protein interfaces (see section 3.1, section 3.2). Second, the reliability of scoring metrics for AI-generated protein complex models is systematically evaluated, resulting in the C2Qscore for combined quality assessment (see section 3.3). Lastly, the analytical scope is extended to protein-nucleic acid and nucleic acid-only interfaces through the development and large-scale application of the ChimeraX plug-in and command-line tool ProNA3D (see section 3.4, section 3.5).

3.1 PICKLUSTER: A Protein-Interface Clustering and Analysis Plug-in for UCSF ChimeraX

To demonstrate the functionality of PICKLUSTER, the ChimeraX plug-in was applied to identify and characterize the protein-protein interface in a heterotrimeric complex from *Epstein-Barr virus* (EBV), modeled using ColabFold (Figure A.1). The implementation of the tool is described in section 6.1. The complex from EBV comprised the proteins BGRF1/BDRF1 (hereafter BDRF1), BFRF1A, and BALF3, which are the EBV homologs of the alphaherpesvirus terminase subunits pUL15, pUL33, and pUL28, respectively (Arakawa et al. 2025). Together, these three proteins constitute the terminase complex, which is responsible for the cleavage of viral DNA concatemers and viral genome packaging (Beard et al. 2002). The analysis specifically focused on the interaction between BDRF1 and BFRF1A (Figure 3.1A). The results presented in this section are based on Genz et al. (2023).

Using the default settings of the heavy atom distance cutoff, d_{HA} , and the clustering distance cutoff, d_C , of 5.0 Å, PICKLUSTER identified three distinct sub-interfaces within the BDRF1-BFRF1A interface (Figure 3.1A). As described in subsection 6.1.1, sub-interface information can be exported as an SCF file, enabling visualization and color-coded annotation of interface residues within the ChimeraX sequence viewer. This feature supports sequence-based downstream analyses of interface regions, such as the evaluation of residue conservation or the identification of putative functional hotspots (Figure 3.1B).

Beyond structural clustering, PICKLUSTER supports pLDDT-based coloring of interface residues (see subsection 6.1.3). Application of this feature to the BDRF1-BFRF1A interface indicated that sub-interface 2 exhibited notably high confidence (Figure 3.1C). In agreement with this observation, the computed ipLDDT value for sub-interface 2

was 90.69, while the corresponding iPAE value was 2.26 Å, indicating low error and highlighting the high quality of the predicted interface (Genz et al. 2025a).

For a more detailed structural assessment, the atomic contact visualization feature of PICKLUSTER was employed (Figure 3.1D). Analysis of interatomic distances and chemical properties of contacting residues suggested a salt bridge between Asp43 (BFRF1A) and Arg67 (BDRF1), as well as a hydrogen bond between Arg67 (BDRF1) and His54 (BFRF1A). These interactions indicated that sub-interface 2 represented a functionally significant sub-interface within the BDRF1-BFRF1A complex.

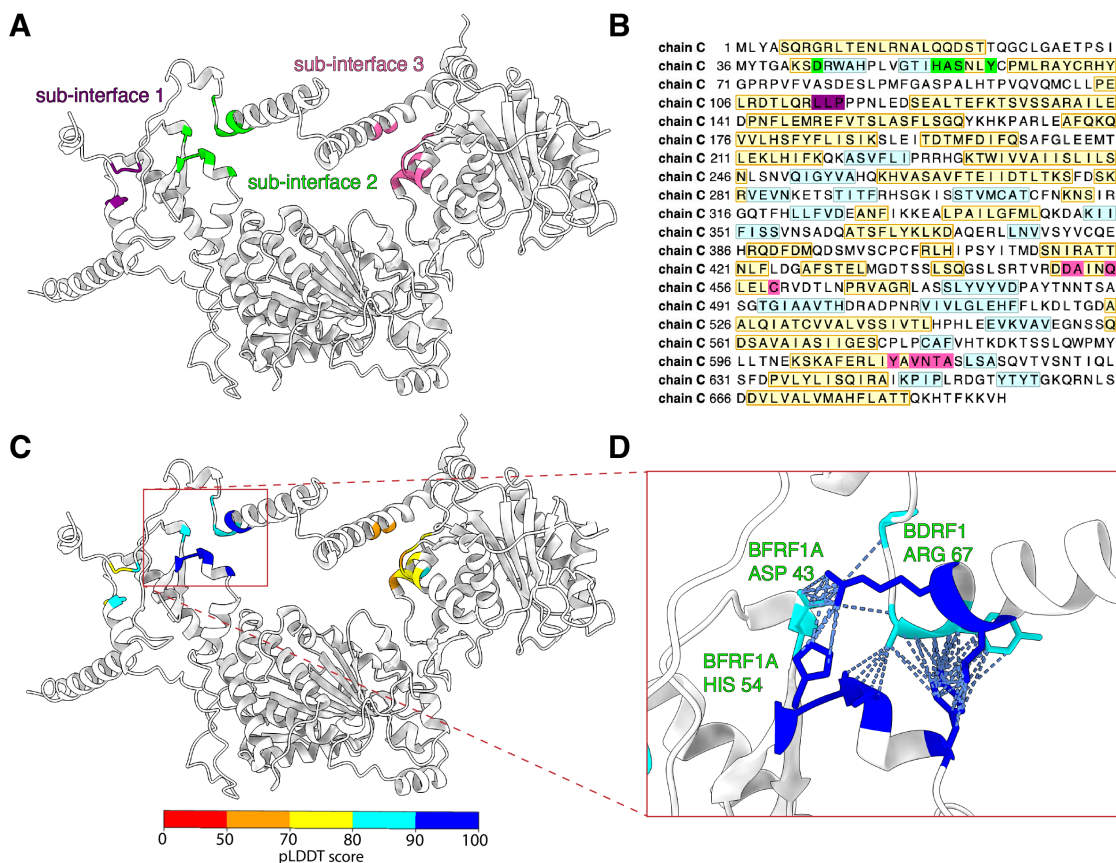


Figure 3.1: PICKLUSTER applied to the dimeric complex of BDRF1 and BFRF1A from *Epstein-Barr virus* (EBV). (A) Visualization of the BDRF1-BFRF1A interaction. Interface residues are clustered into three sub-interfaces, which are shown in purple, green, and pink, based on a distance-based clustering threshold of 5.0 Å. (B) Sequence representation of chain C, where residues are colored according to their corresponding sub-interfaces. (C) The sub-interfaces depicted in (A), colored according to their pLDDT scores. Red indicates low-confidence regions, while blue marks high-confidence areas. (D) Close-up view of sub-interface 2 from panel (A), highlighting the atomic-level interactions. Residues that potentially form specific intermolecular interactions are labeled. Figure adapted from Genz et al. (2023).

As of May 2026, PICKLUSTER as a plug-in has been downloaded 3,371 times from the ChimeraX Toolshed and cited in 12 independent (excluding self-citations) publications, comprising 7 peer-reviewed articles (Akuh et al. 2025; Dan-Dukor et al. 2025; Maurakis et al. 2025; Reguei and Murail 2025; X. Zhou et al. 2025; Machulin et al. 2026; Punga et al. 2026) and 4 preprints (Álvarez-Salmoral et al. 2024; Roddan et al. 2024; Akwelle et al. 2025; Vetrano et al. 2026), as well as 1 undergraduate research journal article (Mohan et al. 2025).

3.2 Benchmarking PICKLUSTER on Large-Scale Protein Complex Data

In addition to the ChimeraX plug-in implementation, PICKLUSTER is available as a command-line tool. To identify structural characteristics of protein-protein interfaces and assess its performance on large datasets, the command-line version was applied to a curated dataset of protein complexes obtained from the PDB. After filtering for protein-only complexes with a resolution of ≤ 3.0 Å and reducing sequence redundancy, 19,821 PDB entries were retained (see subsection 6.1.4). Using default parameters ($d_{HA} = 5.0$ Å, $d_C = 5.0$ Å, subsection 6.1.2), PICKLUSTER identified 29,288 protein-protein interfaces, which were subdivided into a total of 49,321 sub-interfaces.

Analysis of sub-interface sizes, defined as the total number of residues in the sub-interface, revealed a median of 16 residues (mean: 29.00, standard deviation (σ) = 37.05, $n = 49,321$). The distribution showed a trend toward smaller sub-interfaces, with 90% comprising 60 or fewer residues and 9% containing between approximately 60 and 200 residues (Figure 3.2A).

The difference between the median and mean values reflected the presence of a small number of exceptionally large sub-interfaces. Collectively, sub-interface sizes spanned nearly two orders of magnitude (on a log scale), reflecting the broad structural diversity of protein-protein interfaces represented in the PDB. The largest sub-interface in the dataset comprised 989 residues (both chains combined), observed in the TOC-TIC supercomplex from *Chlamydomonas reinhardtii* (PDB ID: 7VCF, Figure 3.2B).

The distribution of sub-interfaces per interface was similarly right-skewed (Figure 3.2C). Approximately 60% of interfaces consisted of a single, continuous interface (median: 1, mean: 1.68, $\sigma = 1.14$, $n = 29,288$), indicating that spatial separation was not the predominant organization of most protein-protein interfaces in the dataset. However, a substantial fraction of interfaces was subdivided into two or three sub-interfaces, while interfaces comprising five or more sub-interfaces were infrequent. The highest number of sub-interfaces observed between any two chains was 18, identified in a synthetic construct comprising two three-helix bundles (PDB ID: 4TQL, Figure 3.2D). Considering the number of sub-interfaces summed across the entire complex, the highest count was 4,844, observed in the phycobilisome structure from *Porphyridium purpureum* (PDB ID: 6KGX), a hetero-706-mer assembly.

Overall, the large-scale analysis using PICKLUSTER provided a comprehensive overview of the size distribution and spatial organization of protein-protein interfaces across experimentally determined structures in the structural proteome.

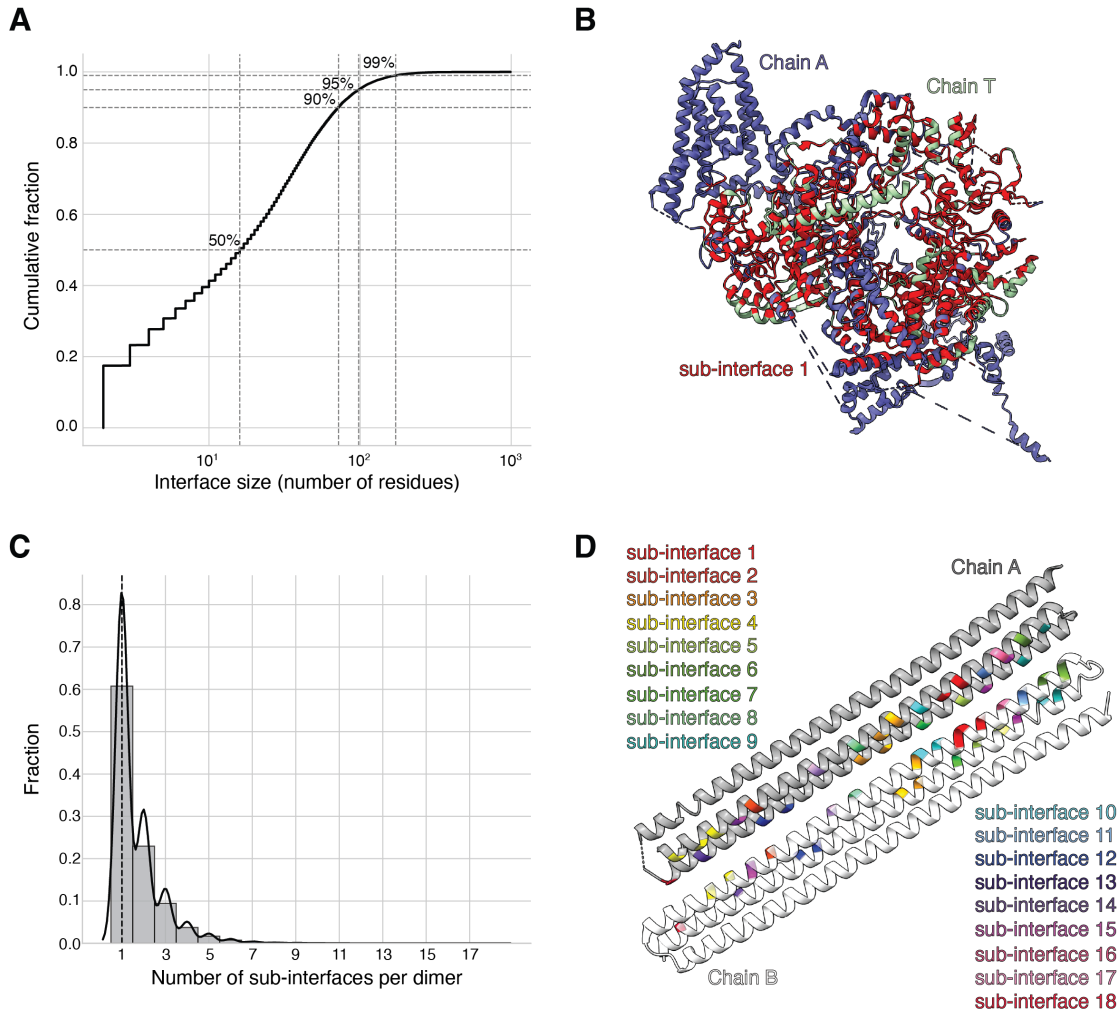


Figure 3.2: Analysis of a large-scale protein complex dataset using PICKLUSTER ($n = 19,821$ PDB entries). (A) Cumulative distribution of sub-interface sizes ($n = 49,321$ sub-interfaces). The x-axis is shown on a logarithmic scale. Dashed lines indicate the 50th, 90th, 95th, and 99th percentiles. (B) The largest sub-interface in the dataset, comprising 989 residues (both chains combined), identified in the TOC–TIC supercomplex from *Chlamydomonas reinhardtii* (PDB ID: 7VCF). Interface residues are shown in red; the two chains involved (Chain A and Chain T) are displayed in dark and light blue, respectively. (C) Distribution of the number of sub-interfaces per protein–protein interface ($n = 29,288$ interfaces). The dashed line indicates the median (one sub-interface per interface). The solid curve shows a kernel density estimate. (D) The highest number of sub-interfaces observed between any two chains in the dataset (18 sub-interfaces), identified in a synthetic construct comprising two three-helix bundles (PDB ID: 4TQL, Chain A and Chain B). Each sub-interface is shown in a distinct color.

3.3 Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions

While PICKLUSTER (Genz et al. 2023) enables identifying, clustering, and analyzing protein-protein interfaces, and offers access to various confidence metrics for AlphaFold2 (Jumper et al. 2021; Evans et al. 2022) and AlphaFold3 (Abramson et al. 2024) predictions, determining which scoring metric reflects the actual accuracy of predicted protein complexes most reliably remains a key challenge.

This section presents a systematic evaluation of widely used scoring metrics for assessing the accuracy of heterodimeric protein complex predictions, with an emphasis on the structural quality of the predicted protein-protein interfaces.

The three structure prediction methods, ColabFold with templates (CF-T) (Mirdita et al. 2022), ColabFold without templates (CF-F), and AlphaFold3 (AF3), were evaluated for differences in their predictive accuracy and corresponding score performance (see subsection 3.3.1). Seven scoring metrics commonly applied in the field of structure prediction were selected for analysis: the interface predicted local distance difference test (ipLDDT) (Jumper et al. 2021; Genz et al. 2023), predicted TM-score (pTM) (Jumper et al. 2021), interface pTM (ipTM) (Evans et al. 2022), *model confidence* (Evans et al. 2022), interface predicted aligned error (iPAE) (Evans et al. 2022; Genz et al. 2023), pDockQ2 (Zhu et al. 2023), and VoroIF-GNN (VoroIF) (Olechnovič and Venclovas 2023). The DockQ (Basu and Wallner 2016) was used as a reference metric for comparison and assessing the quality of the examined scores (see subsection 3.3.2, subsection 3.3.3, subsection 3.3.4, subsection 3.3.5).

All individual prediction-based scores were benchmarked on a curated dataset, referred to as the benchmark set, comprising 223 high-resolution heterodimeric protein complexes from the PDB (see subsection 6.2.1). To improve robustness and interpretability beyond single metrics, ipLDDT, iPAE, pTM, ipTM, and VoroIF were integrated into a combined score, referred to as C2Qscore, which was trained and optimized on the benchmark set (see subsection 3.3.6).

The performance of all individual and combined scores was further examined using two independent test datasets: (i) structures of heterodimeric complexes determined by X-ray crystallography and (ii) dimers extracted from cryo-EM-derived assemblies of high stoichiometry, thereby assessing score generalizability across experimental conditions and complex sizes (see subsection 3.3.6, subsection 3.3.7). To facilitate accessibility, C2Qscore was integrated as an interactive module within the PICKLUSTER v2.0 ChimeraX plug-in, enabling the evaluation of prediction confidence and interface quality directly within a visualization environment.

The work described in this section is based on the publication (Genz et al. 2025a).

3.3.1 Benchmarking the prediction accuracy of ColabFold with templates, ColabFold without templates, and AlphaFold3

To evaluate the performance of ColabFold and AlphaFold3 in generating accurate structural predictions of protein complexes and assessing the reliability of the analyzed scoring metrics, a systematic benchmark analysis was performed. The analysis was restricted to heterodimers, as previous studies have demonstrated that homomeric interfaces are generally predicted with higher accuracy by AlphaFold2 (Evans et al.

2022; Zhu et al. 2023). Focusing on heterodimeric assemblies, therefore, provided a more challenging and diverse test case for the comparative evaluation.

The initial set of structures downloaded from the PDB yielded 684 protein complexes. A series of filtering steps reduced the dataset to 268 target structures (see subsection 6.2.1). Among these, 6 cases were identified where the biological assembly (BA) annotated in the PDB differed from the asymmetric unit (AU), such that the predictions corresponded to only one of the two configurations evaluated using DockQ (see subsection 6.2.1). This discrepancy led to alignment difficulties between the AU and the predicted models, producing artificially low scores that did not accurately represent prediction quality in these cases. Therefore, only templates in which the annotated dimeric BA equaled the AU were retained. Additionally, 39 protein complexes were withheld as an independent test set, resulting in a final benchmark set of 223 target structures.

Protein complex predictions for the benchmark set were generated using the three prediction methods ColabFold with templates (CF-T), ColabFold without templates (CF-F), and the AlphaFold3 web server (AF3) (see subsection 6.2.2). All ColabFold predictions were performed using three recycles followed by structure relaxation (Hornak et al. 2006), yielding five models per target. AlphaFold3 predictions likewise produced five models for each target. In total, DockQ scores (see subsection 6.2.3) and all individual prediction-based confidence metrics (see subsection 6.2.4) were calculated for 1,115 models in each dataset (CF-T, CF-F, and AF3).

Prediction quality of the models was evaluated using the CAPRI criteria for DockQ (Lensink and Wodak 2013). Predictions were classified as incorrect, acceptable, medium, or high quality using threshold values of 0.23, 0.49, and 0.80, respectively (Table 1.1).

Among the three prediction methods, AlphaFold3 achieved the highest proportion of high-quality models (DockQ > 0.8, 39.8%), with CF-T ranking second (35.2%). In both cases, high-quality models accounted for the largest proportion of predictions (Figure 3.3A). CF-F exhibited the smallest proportion of high-quality predictions (28.9%) with models of medium quality constituting its dominant fraction. Examining predictions classified as incorrect (DockQ < 0.23) showed that AlphaFold3 exhibited the lowest incorrect rate (19.2%), whereas ColabFold-based approaches showed markedly higher rates (30.1% for CF-T and 32.3% for CF-F).

The detailed analysis of the incorrect fraction revealed notable differences in prediction consistency among the prediction methods (Figure 3.3B). Within the AF3 predictions, 91.1% of the targets classified as incorrect had all five predicted models assigned to the incorrect category, whereas this proportion was 81.9% for CF-F and 79.1% for CF-T.

Overall, AlphaFold3 generated the most accurate predictions considering only the high-quality fraction. When acceptable, medium, and high quality models were grouped as correct predictions (as opposed to the incorrect fraction), AlphaFold3 showed superior performance, producing the smallest proportion of incorrect models. It emerged as the most reliable prediction method among the evaluated approaches, followed by CF-T. Furthermore, the analysis indicated that compared to ColabFold, AlphaFold3 produced more consistent predictions with respect to model correctness. This is reflected by the higher frequency of cases in which all five generated models were uniformly classified as incorrect, in combination with the lowest proportion of cases in which only a single model was incorrect.

To assess the agreement between the prediction methods, the consensus of prediction quality across the three prediction sets (CF-T, CF-F, and AF3) was

examined by comparison of models of the same rank (Figure 3.3C(i), C(ii), Figure A.2). When all correct models (acceptable, medium, and high quality models based on CAPRI criteria) were considered, the three methods agreed in 78.4% of cases (Figure 3.3C(i)). AlphaFold3 produced the largest fraction of uniquely correct predictions (12.9%), suggesting that using AlphaFold3 may increase the likelihood of generating reliable protein complex models.

When focusing exclusively on high-quality predictions (based on CAPRI criteria), agreement across all methods decreased to 57.1% (Figure A.2). In this subset, AlphaFold3 showed the highest fraction of unique high-quality models (16.8%). Conversely, CF-F displayed the largest fraction of uniquely incorrect predictions (10.4%) based on CAPRI criteria (Figure 3.3C(ii)), consistent with its overall performance for the benchmark set (Figure 3.3A). The lowest level of agreement was observed for predictions classified as of acceptable quality (14.0%) (Figure A.2), suggesting that predictions close to the incorrect threshold (DockQ < 0.23) may be improved by combining the results of multiple prediction methods.

Finally, factors contributing to consistent prediction failure or success were investigated by comparing interface size and flexibility between targets that were mispredicted by all methods (classified as incorrect) and those consistently predicted with high quality. The mispredicted group displayed significantly higher interface flexibility compared to the high-quality group, as quantified by the average B-factor of interface residues (mean: 48.30 ± 20.27 vs. 37.55 ± 27.49) (Figure 3.3D). Interface size also differed substantially between the two groups. Consistently mispredicted targets were characterized by smaller interfaces, reflected in a lower maximum in the kernel density estimate of the interface size distribution (incorrect: 32.75, high quality: 60.43) (Figure 3.3E). Both differences were statistically significant based on the Mann-Whitney U test (see subsection 6.3.11, interface size $p = 4.33 \times 10^{-6}$, B-factor $p = 0.003$), indicating that larger and less flexible interfaces are more likely to result in high-quality structure predictions.

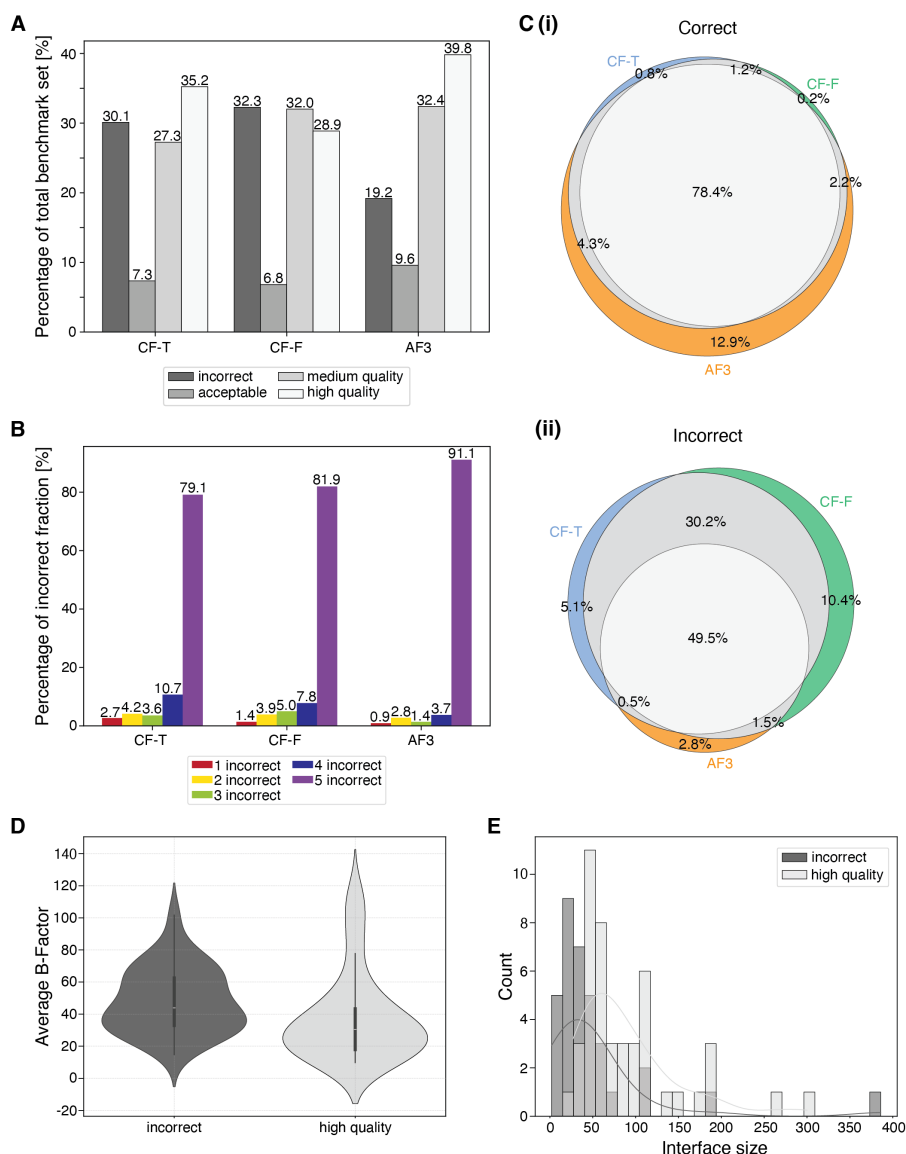


Figure 3.3: Distribution of quality scores, overlap of DockQ classifications (five categories), and detailed analysis of targets across models predicted by CF-T, CF-F, and AF3. (A) Distribution of prediction quality across the three prediction sets based on DockQ classification. **(B)** Detailed breakdown of the incorrect category from (A), showing the number of cases in which all five models of a given template are classified as incorrect, followed by cases with four, three, two, or only one incorrect model. **(C(i))** Venn diagram illustrating the overlap of correctly predicted models (combining acceptable, medium, and high quality). Predictions unique to CF-T are shown in blue, CF-F in green, and AF3 in orange. Overlap between the two methods is indicated in gray, while agreement among all three is shown in light gray. **(C(ii))** Venn diagram of incorrectly predicted models, with color coding as in (C(i)). **(D)** Distribution of the mean interface B-factor for targets that are consistently mispredicted compared to those consistently predicted with high quality across all three methods. **(E)** Distribution of interface sizes for targets that are consistently mispredicted versus those consistently predicted with high quality across all three methods, including kernel density estimates. Adapted from Genz et al. (2025a).

3.3.2 Influence of the heavy atom distance parameter for interface calculation

The identification of protein-protein interfaces, and therefore the calculation of interface-specific confidence measures such as iPAE and ipLDDT, depends on the definition of interacting residues using a d_{HA} threshold. Because the detected interface can vary depending on the selected cutoff, the influence of this parameter on the resulting scores was systematically evaluated. The analysis was performed on the CF-T predictions. ipLDDT and iPAE values were calculated using d_{HA} cutoffs ranging from 5.0 to 9.0 Å in 1.0 Å increments. Across this range, no substantial differences were observed in the resulting ipLDDT and iPAE values (Figure A.3), indicating that these metrics are largely insensitive to moderate variations in the distance threshold. Based on these results, a d_{HA} cutoff of 5.0 Å was selected for the definition of the interface and subsequent score calculations. This choice was consistent with the distance cutoff used by AlphaFold2 for defining interfaces (Jumper et al. 2021).

3.3.3 Distributions and correlation of individual prediction-based scores

To examine the ability of the individual confidence scores (ipLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF) to reflect the quality of the predicted protein complex models, the unnormalized scores were calculated for each model in CF-T, CF-F, and AF3 (see Table A.1, and subsection 6.2.4). To investigate the relationships among these metrics, their empirical score distributions as well as pairwise correlations were analyzed (Figure 3.4, Figure A.4, Figure A.5). Across all prediction sets, the score distributions were generally shifted toward higher values (or lower values in the case of iPAE), indicating an overall high quality of the predicted models, in agreement with the results described in subsection 3.3.1.

Analysis of the empirical distributions (histograms shown in the diagonal panels of Figure 3.4) revealed substantial differences between the metrics. In particular, ipTM and *model confidence* displayed bimodal distributions with their global maxima at high score values. In contrast, ipLDDT, VoroIF, and pTM showed right-skewed distributions, whereas iPAE exhibited a left-skewed pattern. This behavior likely reflected the definition of iPAE, where larger values correspond to higher predicted error and therefore lower model quality, opposite to the interpretation of the other examined metrics. The pDockQ2 score also exhibited a bimodal distribution. However, in contrast to ipTM and *model confidence*, the pDockQ2 distribution exhibited its global maximum at low score values, indicating that a large fraction of models were assigned very low predicted interface quality.

Among all pairwise relationships evaluated using Pearson correlation coefficients (PCC) (Figure 3.4, Figure A.4, Figure A.5, see subsection 6.2.6), *model confidence* and ipTM exhibited the strongest correlation across all three prediction sets. This was expected, as ipTM represents the main component of the *model confidence* metric. In contrast, pDockQ2 showed the weakest correlations with the other interface-specific metrics. This pattern likely reflected the pDockQ2 definition, which applies a sigmoid transformation to a combination of interface-specific pLDDT and iPAE (Zhu et al. 2023). This results in compression of score values towards both extremes of the sigmoid, reducing the overall score variance.

Overall, the distributions and correlations of the individual prediction-based scores were consistent with the observation that all three prediction sets contained a high proportion of high-quality models. Importantly, these trends were persistent across CF-T, CF-F, and AF3 predictions (Figure 3.4, Figure A.4, Figure A.5), indicating that the observed relationships are largely independent of the prediction method used.

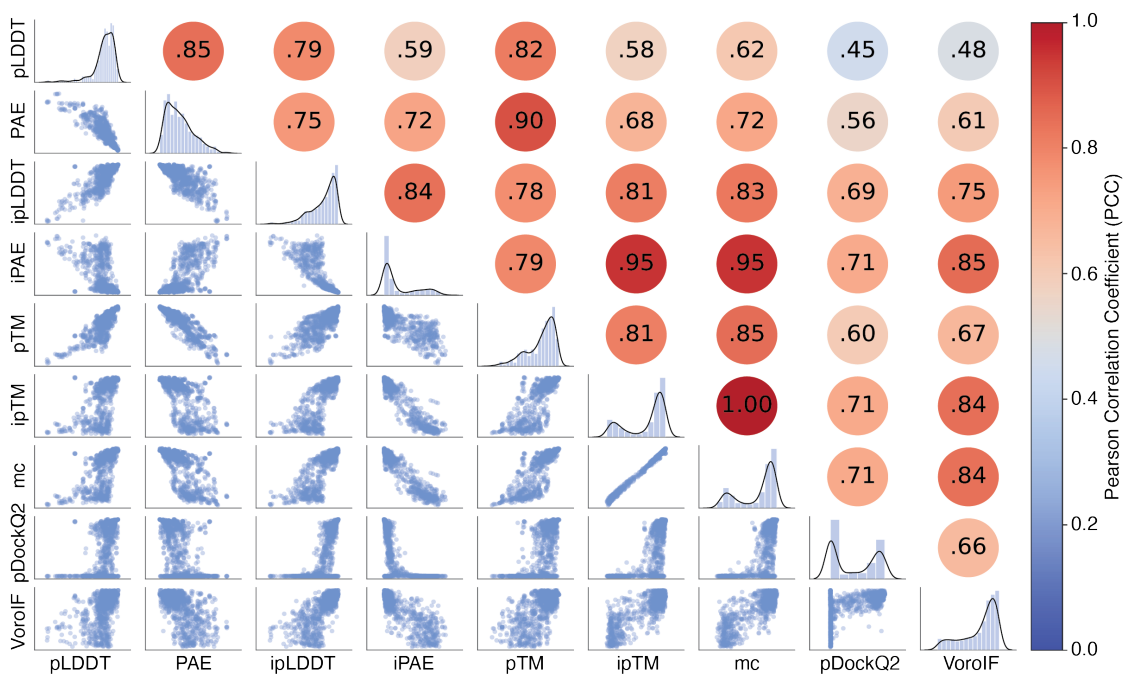


Figure 3.4: Distributions and correlations of model quality scores for the CF-T predictions. The evaluated metrics include pLDDT, PAE, ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VorolIF. The diagonal panels show the distribution of each score across the benchmark set. Pairwise relationships between scores are shown in the lower-left panels. The upper-right panels report the corresponding Pearson correlation coefficients (PCCs), with red indicating stronger correlations and blue indicating weaker correlations. Adapted from Genz et al. (2025a).

3.3.4 Correlation of the individual prediction-based scores with DockQ

To evaluate the ability of the individual prediction-based scores to reflect the structural quality of predicted protein complexes, the PCC between each score and DockQ, which was used as a reference metric, was calculated for all three prediction sets (see subsection 6.2.6). Mean PCCs, averaged across all individual scores, were 0.70, 0.72, and 0.64 for CF-T, CF-F, and AF3, respectively. A Wilcoxon signed-rank test (see subsection 6.2.10) indicated that the differences between each of the ColabFold prediction sets and AF3 were statistically significant ($p = 0.004$), whereas no significant difference was observed between CF-T and CF-F. Notably, despite the comparatively lower proportion of high-quality models in the CF-F predictions (Figure 3.3A), the scores demonstrated the strongest agreement with DockQ for this prediction method (Figure 3.5A). These results suggested that the individual scores reflect the model quality

more reliably for structures generated by ColabFold without templates than for those predicted by AlphaFold3.

Across all three prediction sets, ipTM and *model confidence* consistently showed the highest PCCs with DockQ (Figure 3.5A, Figure A.6A), identifying them as the most reliable individual metrics for the assessment of predicted heterodimeric protein complexes, independent of the prediction method applied. The interface-specific scores ipLDDT and iPAE exhibited higher PCCs with DockQ than their global counterparts, pLDDT and PAE, consistent with the results from subsection 3.3.3. In contrast, pDockQ2 showed the weakest correlations with DockQ across all prediction sets, likely due in part to the accumulation of near-zero values also described in subsection 3.3.3.

Representative structural examples highlighting both agreement and disagreement between ipTM and pDockQ2 assessments are presented in Figure 3.5B. In the first example, ipTM (0.82) correctly assigned high confidence to a high-quality prediction (PDB ID: 9H8S, DockQ: 0.83), whereas pDockQ2 (0.17) substantially underestimated the model quality (Figure 3.5B(i)). Conversely, a case in which both metrics correctly agreed on low model quality is shown in Figure 3.5B(ii) (PDB ID: 8B2R, ipTM: 0.16, pDockQ2: 0.01, DockQ: 0.19).

Figure 3.5B(iii) illustrates a case in which both scoring metrics incorrectly assessed model quality. Despite representing an incorrect prediction (PDB ID: 7PPZ, DockQ = 0.13), the model exhibited high ipTM (0.85) and pDockQ2 (0.81) values. This discrepancy arose from an incorrect domain orientation that led to a mispredicted interface. Figure 3.5B(iv) illustrates a case in which both scoring metrics correctly indicated a high-quality prediction (PDB ID: 7PAY, ipTM: 0.91, pDockQ2: 0.82, DockQ: 0.94).

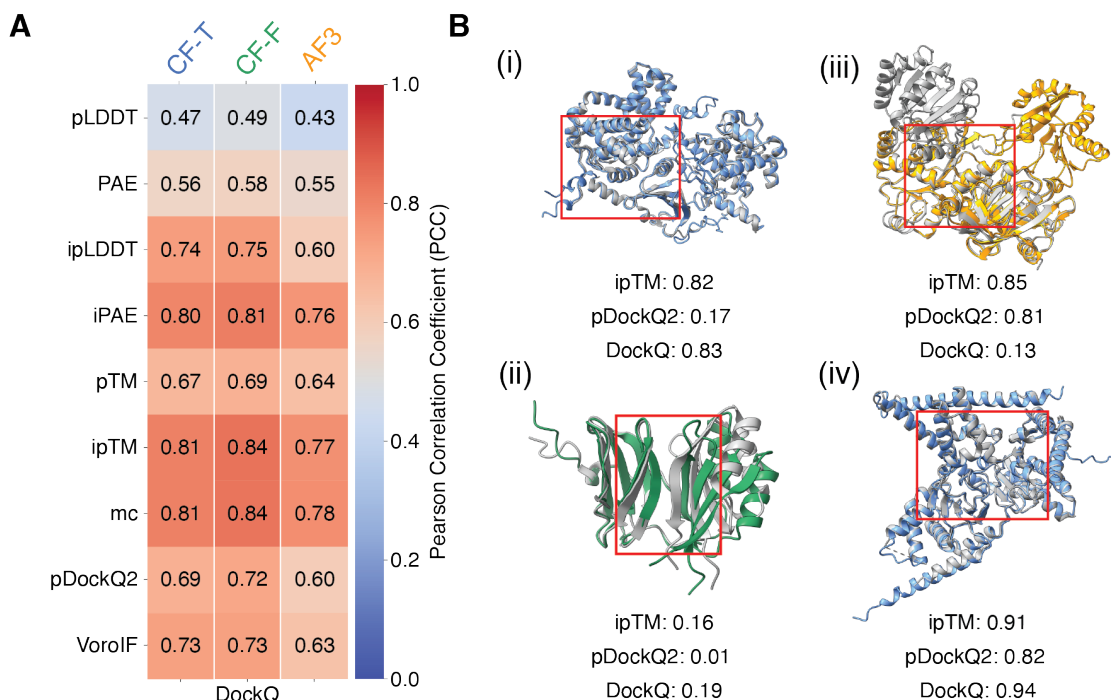


Figure 3.5: Pearson correlation coefficients of individual scores with DockQ and representative structural predictions. (A) PCC of individual scores (pLDDT, PAE, ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VoroIF) with DockQ for CF-T, CF-F, and AF3. A color gradient from blue to red indicates low to high PCC. (B(i)) CF-T prediction (blue) and target structure (PDB ID: 9H8S, gray), ipTM: 0.82, pDockQ2: 0.17, DockQ: 0.83. (B(ii)) CF-F prediction (green) and target structure (PDB ID: 8B2R, gray), ipTM: 0.16, pDockQ2: 0.01, DockQ: 0.19. (B(iii)) AF3 prediction (orange) and target structure (PDB ID: 7PPZ, gray), ipTM: 0.85, pDockQ2: 0.81, DockQ: 0.13. (B(iv)) CF-T prediction (blue) and target structure (PDB ID: 7PAY, gray), ipTM: 0.91, pDockQ2: 0.82, DockQ: 0.94. In B(i)-(iv), the interface region is indicated by a red box. Adapted from Genz et al. (2025a).

3.3.5 Determinants of high- and low-quality models

The individual prediction-based scoring metrics provide absolute values without established thresholds, which limits their direct application to the binary classification of a model as correct or incorrect. To address this limitation, receiver operating characteristic (ROC) curve analysis was performed for each score across all three prediction sets (see subsection 6.2.7). This enabled the determination of optimal classification cutoffs (Figure 3.6A), following the approach of (Melo and Sali 2007).

For this analysis, models assigned to the incorrect DockQ category (DockQ < 0.23) were considered incorrect, while models of acceptable, medium, or high quality according to CAPRI criteria were classified as correct predictions. The reported cutoffs apply exclusively to protein complexes of the type represented in the benchmark dataset and do not apply to antibody-antigen interactions, which were excluded from the benchmark prior to the analysis (see subsection 6.2.1). In addition, the area under the curve (AUC) (Figure 3.6B), along with the accuracy and the precision (Table A.2), were computed to evaluate the efficacy of the selected thresholds in discriminating between incorrect and correct models as defined by DockQ (see subsection 6.2.8).

All score-specific cutoff values determined for each prediction set are summarized in Figure 3.6A. Examination of the derived cutoffs revealed that the optimal thresholds for pDockQ2 were unexpectedly low across all three prediction sets (CF-T: 0.15, CF-F: 0.05, AF3: 0.05), given the scoring range of 0 to 1. The AUC values and classification accuracy were highest for ipTM and *model confidence* across all prediction sets (Figure 3.6B, Supplementary Table S2), consistent with the results of subsection 3.3.4. AUC and accuracy values for the AF3 predictions were lower than those obtained for CF-T and CF-F, reflecting the comparatively weaker agreement between the individual metrics and DockQ for AlphaFold3 predictions, as described in subsection 3.3.4.

The AlphaFold guidelines define an ipTM range of 0.6 to 0.8 as a gray zone in which the quality of a prediction cannot be reliably assessed (Gomez and Kovalevskiy 2025a). The ROC-derived cutoffs were therefore evaluated specifically for models with ipTM values within this range. Applying the dataset-specific thresholds (CF-T: 0.65, CF-F: 0.65, AF3: 0.70) to this subset yielded classification accuracies of 0.82, 0.86, and 0.87 for CF-T, CF-F, and AF3, respectively (Table A.3), demonstrating that the determined cutoffs enabled reliable binary classification even within the ambiguous gray zone.

Representative structural examples from this subset are shown in Figure 3.6C. In two cases, the predicted models showed ipTM scores below the respective dataset-specific cutoffs and DockQ values below 0.23 and were therefore correctly classified as incorrect. These included a CF-T prediction for the target PDB ID 7XYQ (ipTM: 0.62, DockQ: 0.06) and an AF3 prediction for the target PDB ID 7FGM (ipTM = 0.66, DockQ = 0.18). In contrast, the CF-F prediction for the target PDB ID 7EGS exhibited an ipTM score of 0.74, exceeding the CF-F ipTM cutoff of 0.65, and was consequently correctly classified as a high-quality prediction (DockQ: 0.86).

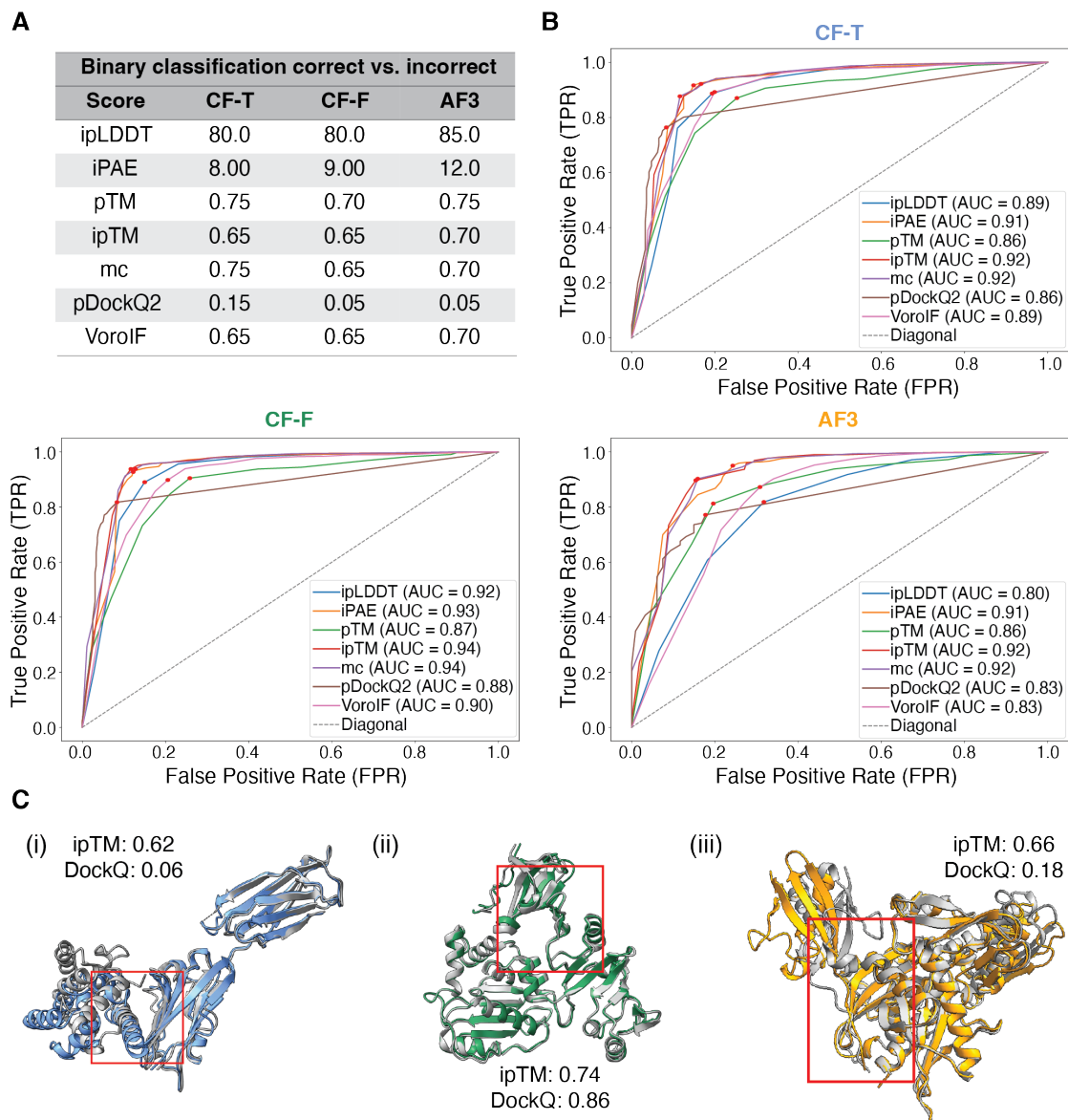


Figure 3.6: Optimal classification cutoffs, ROC curves, and representative gray zone predictions. (A) Optimal thresholds for binary classification of correct versus incorrect models, derived by ROC analysis for CF-T, CF-F, and AF3. (B) ROC curves and AUC values for all individual scores (ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VoroIF) across CF-T, CF-F, and AF3. (C) Representative structural predictions with ipTM values in the gray zone. (C(i)) CF-T prediction and target structure (PDB ID: 7XYQ, gray), ipTM: 0.62, DockQ: 0.06. (C(ii)) CF-F prediction and target structure (PDB ID: 7EGS, gray), ipTM: 0.74, DockQ: 0.86. (C(iii)) AF3 prediction and target structure (PDB ID: 7FGM, gray), ipTM: 0.66, DockQ: 0.18. Interface regions are indicated by a red box. Adapted from Genz et al. (2025a).

3.3.6 Combined scoring and testing

The individual prediction-based scores are informative, but each reflects only a partial aspect of the interface quality. To enable a more robust model assessment beyond single metrics, a combined scoring approach was developed (see subsection 6.2.5). As *model confidence* and pDockQ2 are themselves derived from combinations of other evaluated scoring metrics, they were excluded from this approach. The remaining five metrics, ipLDDT, iPAE, pTM, ipTM, and VoroIF, were selected as input variables.

Among the tested combination approaches, a linear regression-based model applied to all five input scores consistently achieved the highest PCCs with DockQ across CF-T, CF-F, and AF3 (Table A.4). This combined score is referred to as C2Qscore (Combined Complex Quality score). The regression weights, biases, and maxima are provided in Tables A.5 to A.7. The full set of evaluated combinations is summarized in Table A.4. A Wilcoxon signed-rank test (see subsection 6.2.10) indicated that C2Qscore provided a statistically significant improvement in model classification compared to an unweighted combination of the individual scores (Table A.8). C2Qscore ranged from -0.45 to 0.88 for CF-T, from -0.37 to 0.87 for CF-F, and from -0.33 to 0.93 for AF3. ROC-based analysis was used to determine an optimal classification cutoff for C2Qscore (Table A.9). Cutoffs of 0.55 (AUC: 0.92), 0.48 (AUC: 0.94), and 0.52 (AUC: 0.92) were determined for CF-T, CF-F, and AF3, respectively.

The performance of C2Qscore was further evaluated on an independent X-ray crystallography test set (resolution ≤ 3.0 Å) comprising 39 heterodimeric protein complexes, which had been randomly selected and excluded from the benchmark set during dataset construction (see subsection 6.2.1). DockQ and all prediction-based scores, including the combined scores, were calculated for 195 models in each of the three prediction sets. The classification cutoffs established in subsection 3.3.5 and in the present subsection were applied to calculate accuracy, precision, recall, and specificity (Figure 3.7). For CF-T, C2Qscore achieved the same classification accuracy as the unweighted combined score. For CF-F and AF3, C2Qscore showed slight improvements relative to the unweighted combination (Figure 3.7). When compared to the individual scores, C2Qscore performed equally well or better than nearly all metrics across all three prediction sets (Figure A.7). Exceptions were observed for iPAE and ipTM in the CF-T predictions, where precision and specificity were equal to those of C2Qscore, but accuracy and recall were slightly higher for the individual scores.

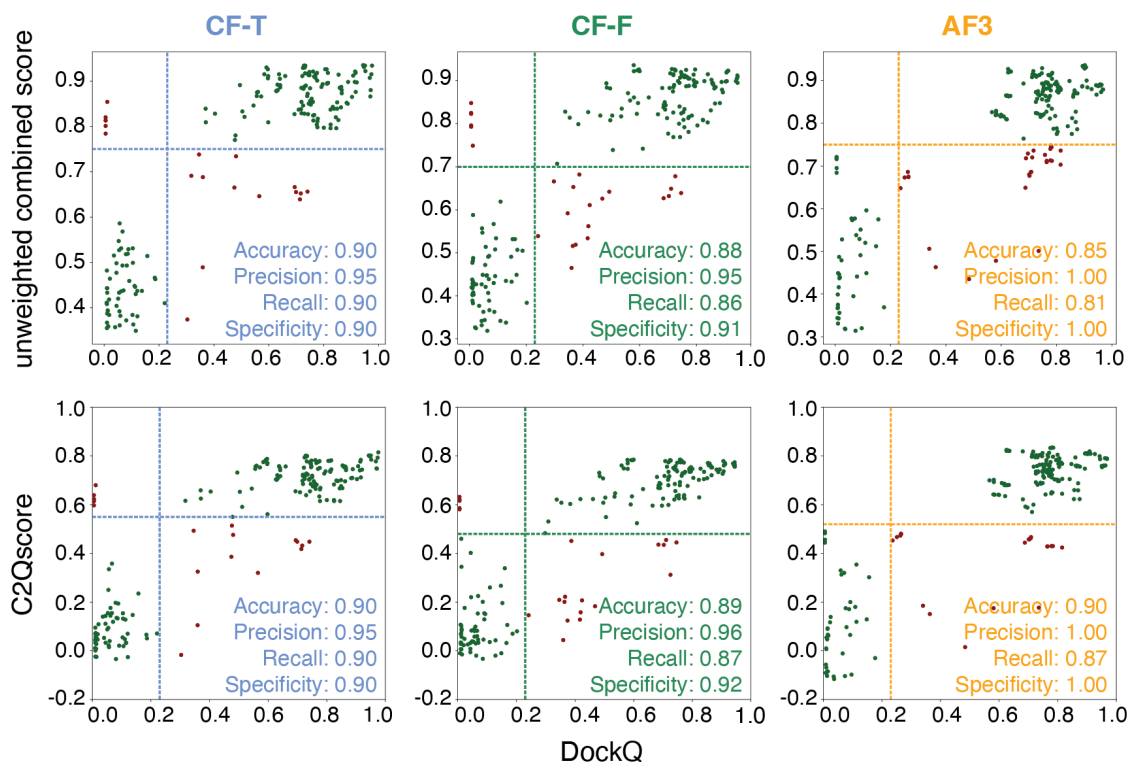


Figure 3.7: Unweighted combined score and C2Qscore evaluated on the X-ray crystallography test set for CF-T, CF-F, and AF3. Scatter plots of the unweighted combined score (top row) and C2Qscore (bottom row) against DockQ. True positives and true negatives are plotted in dark green. False positives and false negatives are displayed in dark red. Classification cutoffs are indicated by dotted lines. Accuracy, precision, recall, and specificity are reported for each dataset. Adapted from Genz et al. (2025a).

3.3.7 Evaluating the scores on dimers from larger assemblies

The performance of the combined scoring metrics was further evaluated on a dataset of dimers extracted from assemblies of higher stoichiometry to test their applicability in a more complex biological context. This dataset comprised 156 dimeric target structures derived from trimeric or higher-order cryo-EM-derived assemblies resolved at a resolution of ≤ 2.3 Å, referred to as the large assembly sets (see section 6.2.1). DockQ and all prediction-based scores, including the combined scores, were calculated for 768 models in the CF-T predictions, 769 in CF-F, and 720 in AF3. The differences in model counts across prediction sets resulted from cases in which score calculation failed, leading to the exclusion of the corresponding models (see section A.3). The classification cutoffs determined in subsection 3.3.5 and subsection 3.3.6 were applied to compute accuracy, precision, recall, and specificity (Figure 3.8A).

Compared to the X-ray crystallography test set, a general decrease in classification accuracy was observed across all prediction methods and scoring approaches, with an increased proportion of false negatives (Figure 3.7, Figure 3.8A, Figure A.8). C2Qscore performed similarly to the unweighted combined score, showing mild improvements for CF-T and AF3, while a slight decrease was observed for CF-F (Figure 3.7).

A key factor contributing to this reduced performance was the presence of multiple possible spatial arrangements of two chains forming a dimer (configurations) within the larger assembly. When two chains constitute a dimer that is part of

a higher-order complex, homologous counterpart chains within the assembly can interact in more than one structurally distinct configuration (Figure 3.8B). Since AlphaFold2 and AlphaFold3 use only the sequence as user input, it is not possible to determine in advance which configuration corresponds to the predicted structure. Consequently, when DockQ is computed against the target configuration, predictions that closely resemble an alternative configuration receive low DockQ scores despite representing accurate models.

A representative example is shown in Figure 3.8C. In this case, the target configuration corresponded to the dimer formed by chains D and I from PDB ID 9ES7. However, the CF-T prediction of this dimer closely resembled an alternative configuration involving chains D and A (homologous to chain I) (DockQ for D-A: 0.71), while the DockQ computed against the target configuration D-I was only 0.01. Despite this, C2Qscore assigned a value of 0.65 to this prediction, reflecting its high structural quality with respect to the alternative configuration.

In a typical prediction scenario, where no ground truth structure is available, the number of possible configurations of a complex is inherently unknown. This makes it essential to evaluate the reliability of scoring metrics when multiple configurations are possible. To investigate this, the large assembly set was used to examine the relationship between DockQ and C2Qscore and the number of configurations per target dimer across the entire large assembly dataset. The analysis was restricted to templates that exhibit more than one dimeric configuration, specifically those with two or three configurations. In addition, only cases in which at least one configuration was classified as correct according to DockQ ($\text{DockQ} \geq 0.23$) were retained. It should be noted that the DockQ values reported in Figure 3.8B are based on the alignment between the prediction and the designated target structure. Consequently, low DockQ scores can arise when a prediction closely matches an alternative configuration but is evaluated against the differing target configuration.

Analysis of the DockQ and C2Qscore distributions for targets with two possible configurations ($n = 58$) in CF-T revealed a bimodal DockQ distribution, with pronounced maxima at very high and very low values (Figure 3.8B). In contrast, the C2Qscore distribution was shifted toward higher score values, with a large fraction of models assigned very high or medium-to-high scores. A similar trend was observed for targets with three possible configurations ($n = 42$). DockQ values were predominantly clustered at low scores, reflecting the presence of multiple possible configurations, while C2Qscore was skewed toward medium-to-high values. Similar patterns were observed for the CF-F and AF3 predictions (Figure A.9).

These findings suggest that, in cases with multiple possible configurations, AlphaFold2 and AlphaFold3 tend to assign lower confidence scores (medium-to-high) to highly accurate predictions. This discrepancy between structural accuracy and assigned confidence may explain the reduced classification accuracy observed for the large assembly dataset compared to the X-ray crystallography test set.

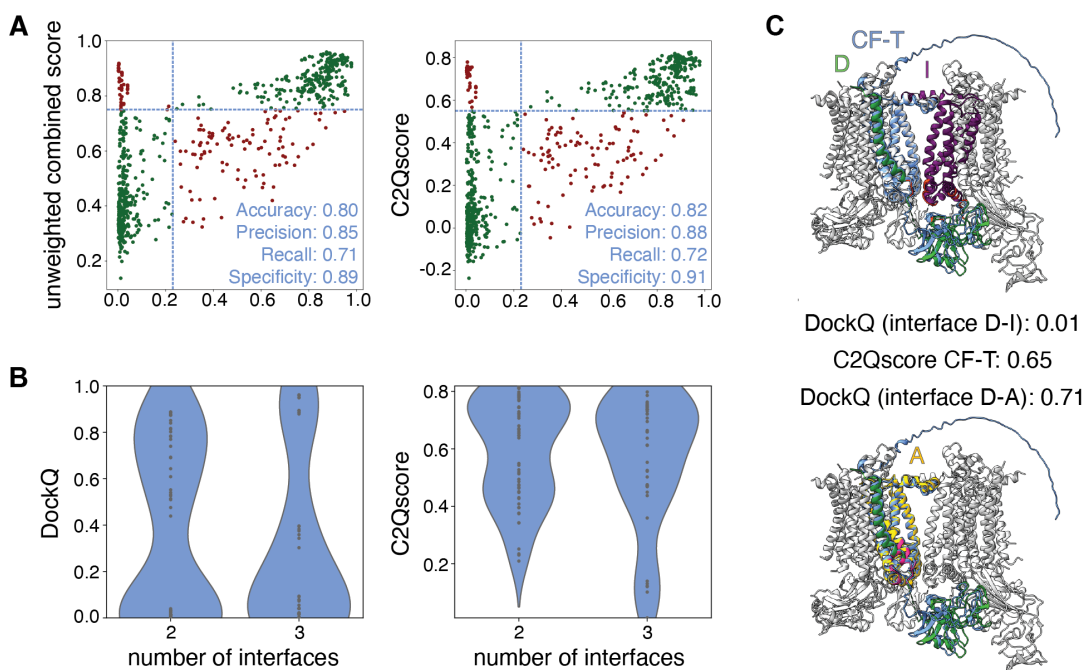


Figure 3.8: Performance of the unweighted combined score and C2Qscore on dimeric interfaces extracted from cryo-EM assemblies of higher stoichiometry (CF-T). (A) Scatter plots showing the relationship between DockQ and the unweighted combined score (top panels) or C2Qscore (bottom panels). True positives and true negatives are highlighted in dark green, whereas false positives and false negatives are shown in dark red. Classification thresholds are indicated by dashed lines. (B) Distributions of DockQ (left) and C2Qscore (right) for cases with two ($n = 58$) or three ($n = 42$) possible dimer configurations. (C) Superposition of a CF-T prediction with the corresponding large assembly (PDB ID: 9ES7). Chains D (green), I (purple), and the homologue of chain I (chain A, gold) are highlighted. The interfaces corresponding to the target configuration (D-I) and an alternative configuration (D-A) are shown, together with their respective DockQ and C2Qscore values. Adapted from Genz et al. (2025a).

C2Qscore and the individual scores, including their classification cutoffs, provide a unified approach for analyzing protein-protein complex predictions, enabling both interface identification and quality assessment of AlphaFold-generated models. However, this approach is designed for protein-only systems and does not account for complexes that involve nucleic acids. Such assemblies exhibit additional structural features and interaction patterns that are not fully captured by methods developed for protein interfaces, highlighting the need for a more general framework.

3.4 ProNA3D: Distance-Based Analysis of Nucleic Acid and Protein Interfaces

This section describes the application of ProNA3D (Figure 3.9), a tool designed for the comprehensive analysis of protein-nucleic acid and nucleic acid-only complexes. While it builds on concepts established in PICKLUSTER, ProNA3D represents a standalone tool with expanded capabilities for interface detection and the identification of spatially separated sub-interfaces for a broader range of macromolecular interfaces.

It combines three-dimensional structural analysis, interface connectivity graphs, and secondary-structure topology plots. Importantly, it supports structure prediction analysis in the context of confidence-based metrics, which is essential for analyzing protein-nucleic acid complexes that often suffer from low confidence in interface regions (see subsection 1.5.2). In addition, ProNA3D enables analysis of cryo-EM density maps via density zoning, facilitating examination of interfaces and flexible regions that often occur in low-resolution regions, particularly in composite maps and RNA-containing complexes. Developed as a plug-in for ChimeraX and as a command-line tool, ProNA3D provides a unified, interactive, and user-friendly platform for analyzing macromolecular interfaces, bridging the gap between structural prediction and functional interpretation. The implementation of ProNA3D is detailed in section 6.3.

In the following, the application of ProNA3D is demonstrated on a range of complex types, including both experimentally resolved structures (X-ray crystallography and cryo-EM) and AlphaFold3-generated models. The available interface-calculation modes are first illustrated using the X-ray structure of an RNA-guided nuclease-RNA complex from *Acanthamoeba polyphaga* mimivirus (see subsection 3.4.1). Further examples comprise the X-ray structure of a dimeric hammerhead ribozyme RNA complex (PDB ID: 2OEU, subsection 3.4.2), a predicted human Fab antibody bound to HIV-1 RNA (target PDB ID: 9C75, subsection 3.4.3), a predicted complex of the Cre recombinase from *Punavirus P1* with its target DNA (target PDB ID: 7RHY, subsection 3.4.4), and the cryo-EM-derived structure of the *Bos taurus* ATP-dependent DNA/RNA helicase DHX36 in complex with human G-quadruplex RNA (PDB ID: 8VV2, EMD-43547, subsection 3.4.5).

This section is based on a study published as a preprint (Genz and Topf 2026). The corresponding manuscript is currently under review for peer-reviewed publication.

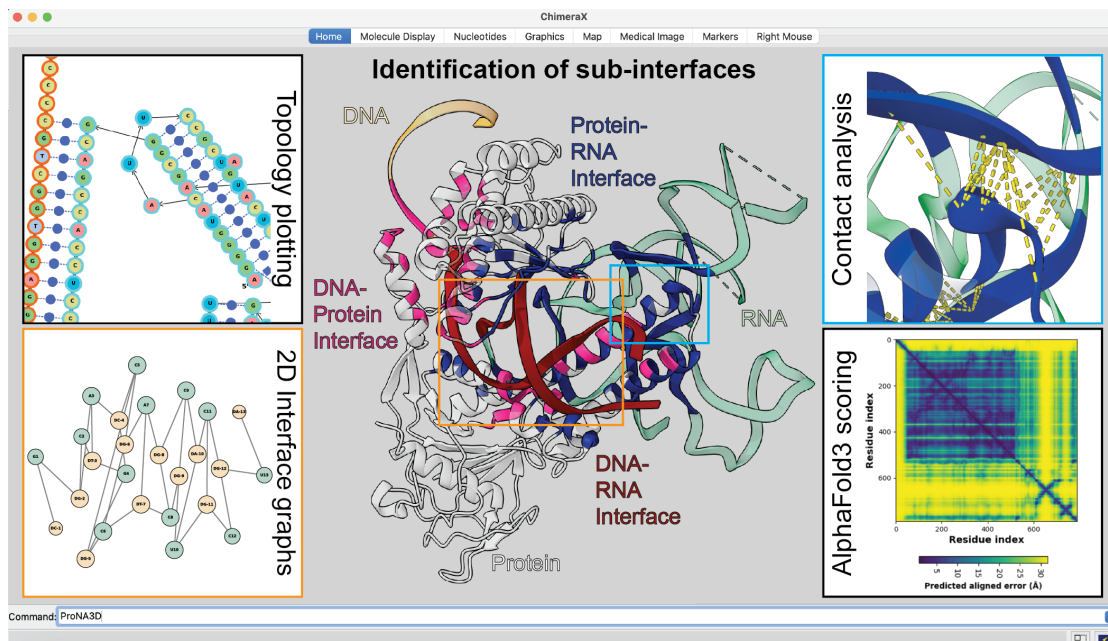


Figure 3.9: Overview of ProNA3D for integrated interface analysis in UCSF ChimeraX. The central panel shows a representative protein-DNA-RNA complex (PDB ID: 9B0L), in which ProNA3D detects and annotates spatially separated sub-interfaces. The panels on the left present interactive two-dimensional views, including a secondary structure topology diagram (upper left) and an interface network representation, where residues and nucleotides are depicted as nodes connected by interaction edges (lower left, orange). The panels on the right display three-dimensional visualizations of interface contacts (upper right, blue) and the integration of structure prediction confidence measures from AlphaFold3, such as the PAE (lower right). Adapted from Genz and Topf (2026).

3.4.1 X-ray structure of a protein-RNA complex: *Acanthamoeba polyphaga* mimivirus RNA guided nuclease-RNA

To compare the three different interface-calculation modes of ProNA3D (see subsection 6.3.1), it was applied to identify the interfaces in the RNA-guided nuclease Fanzor2 (chain P) bound to an RNA scaffold (chain C) and its target DNA (chains A, B, and D). The complex from *Acanthamoeba polyphaga* mimivirus (Schargel et al. 2025) (PDB ID: 9B0L) was included as a target in CASP16 (Figure 3.10A). The default d_{HA} cutoff was applied without division into sub-interfaces. Pairwise analysis of chains A and P (Figure 3.10B) yielded a single continuous interface between the two chains, highlighting how this mode can be used for a detailed examination of specific contacts in the interface.

Applying ProNA3D across the entire Fanzor2 complex identified 6 distinct interfaces: A:B, A:C, A:P, B:P, C:P, and D:P (Figure 3.10C). This global overview enabled rapid identification of interacting components within assemblies of higher stoichiometry.

Calculation of interfaces for chain P against all other chains in the complex revealed that Fanzor2 interacts with all other components (Figure 3.10D). This observation is consistent with the biological role of Fanzor2 as an RNA-guided nuclease and supports the structural stability of the assembly.

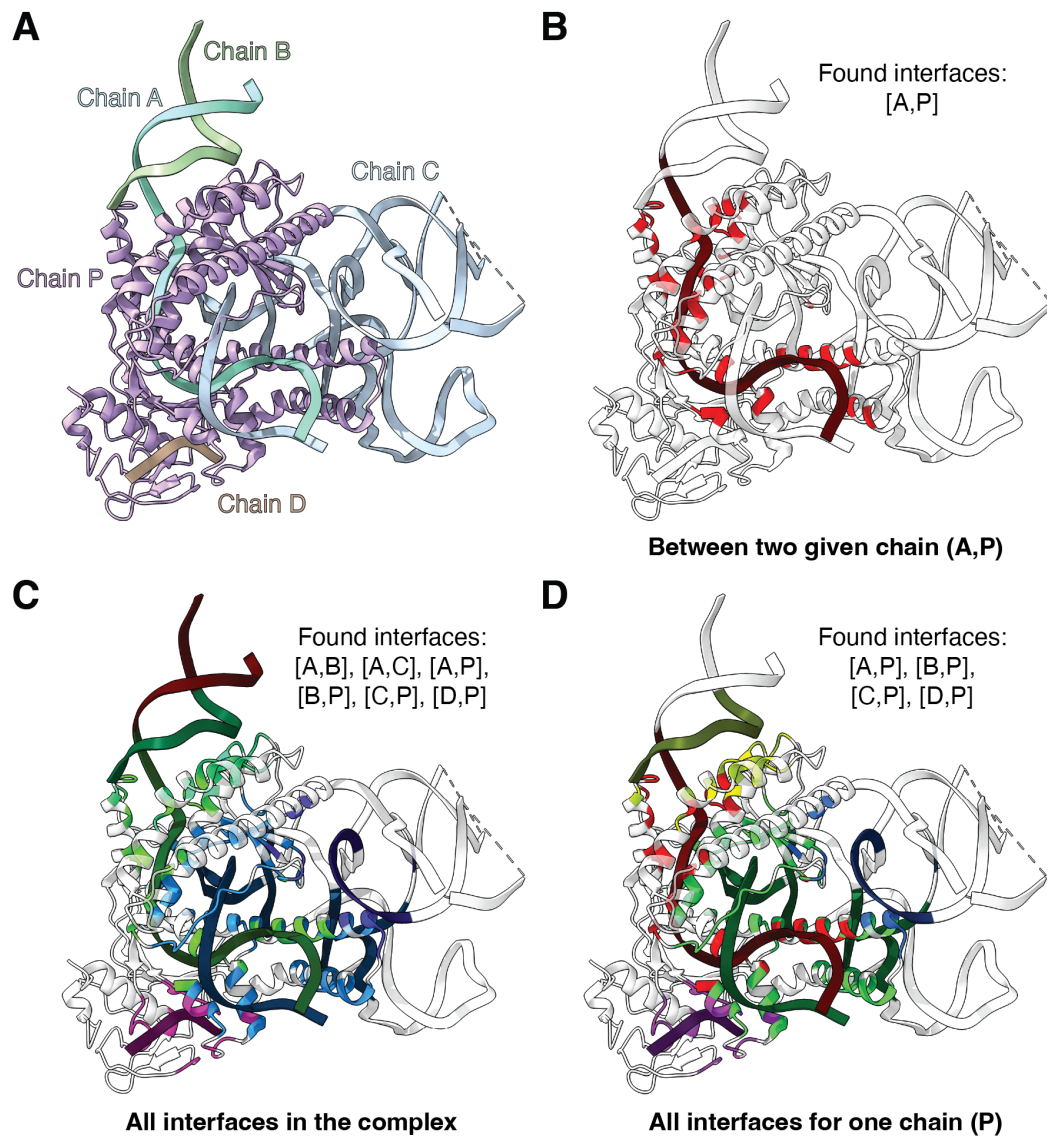


Figure 3.10: Interface modes provided by ProNA3D. Residues participating in interfaces are shown in color, whereas non-interface residues are displayed in gray. **(A)** Overall structure of the RNA-guided nuclease Fanzor2 (chain P) in complex with its RNA scaffold (chain C) and target DNA (chains A, B, and D) from *Acanthamoeba polyphaga* mimivirus (Schargel et al. 2025) (PDB ID: 9B0L). **(B)** Pairwise interface calculation for two user-specified chains (A and P). **(C)** Interface calculation across the entire complex, identifying all interfaces present. **(D)** Interface calculation for all interactions involving a single user-specified chain (P).

3.4.2 X-ray structure of an RNA-only complex: hammerhead ribozyme

To assess the impact of the d_{HA} parameter, ProNA3D was applied to a dimeric hammerhead ribozyme RNA complex (PDB ID: 2OEU) across a range of values. Using a d_{HA} of 3.0 Å (see subsection 6.3.2), five distinct sub-interfaces within the dimeric complex were identified (Figure 3.11A). Increasing the cutoff to 4.0 Å reduced the number of detected sub-interfaces to two, whereas a cutoff of 5.0 Å resulted in a single, continuous interface. These results demonstrated that the choice of the d_{HA} parameter can substantially influence the outcome of the interface calculation.

In addition, the topology of the RNA-RNA complex was analyzed using the Leontis-Westhof base pair classification (see subsection 6.3.3, Figure 3.11B, C). ProNA3D identified a total of 29 base-pair interactions, comprising 9 intra-chain and 20 inter-chain contacts. The resulting topology plot highlighted nucleotide A38 in chain A as a central interaction hub, suggesting potential functional relevance (Figure 3.11B). A38 formed backbone interactions with nucleotides 37 and 39 and engaged in a cis Hoogsteen/Sugar-edge pairing with nucleotide 17 of the same chain. Furthermore, A38 participated in a trans Watson-Crick/Hoogsteen interaction with the 2'-O-methylated nucleotide at position 6 of chain B. This inter-chain contact was part of the interface detected by ProNA3D (Figure 3.11C). Examination of A38 within the three-dimensional structure further supported its structural importance, as it coordinates a Mn^{2+} ion (Figure 3.11D).

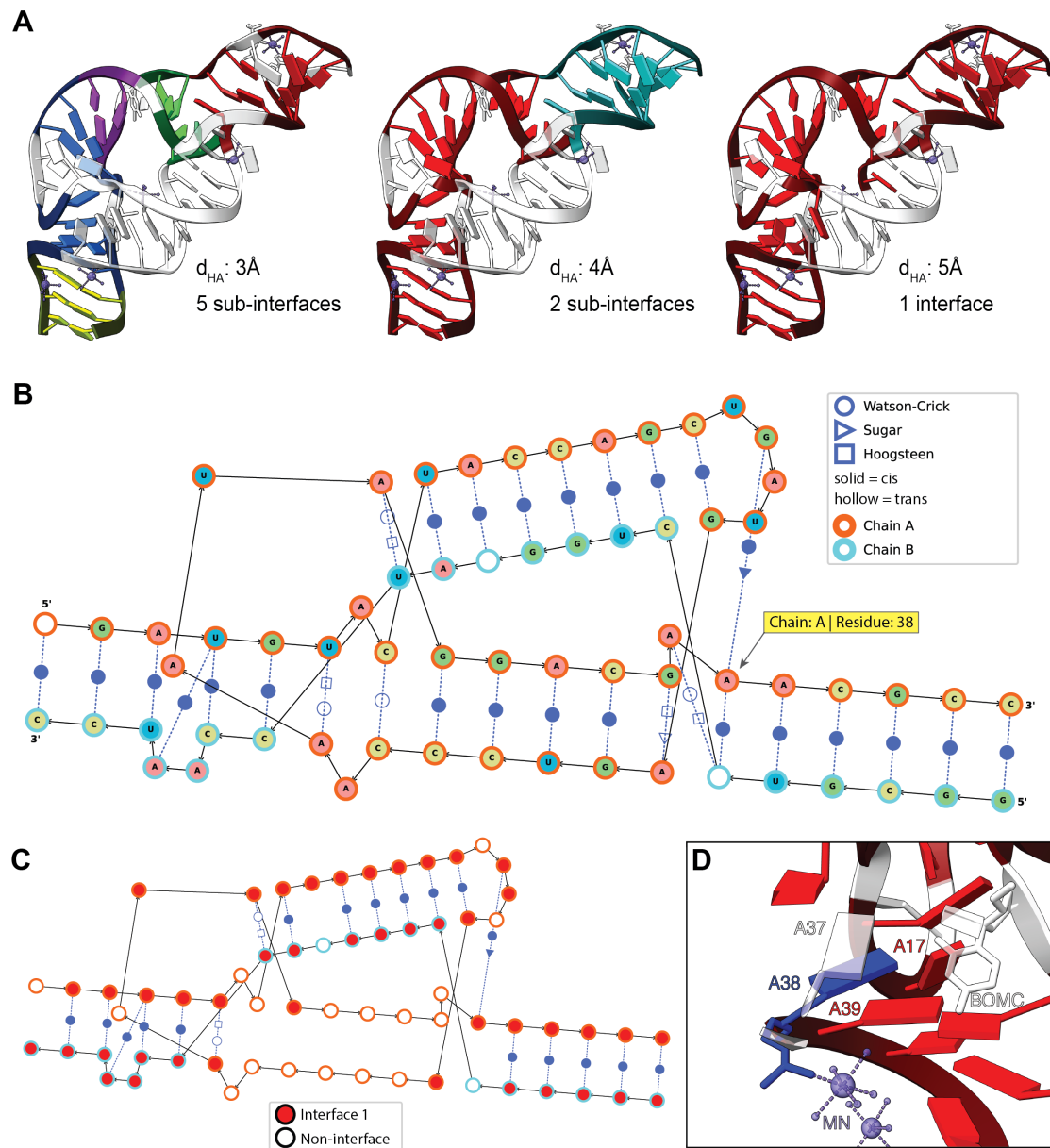


Figure 3.11: Effect of the heavy-atom distance cutoff (d_{HA}) and topological analysis of a dimeric hammerhead ribozyme RNA complex (PDB ID: 2OEU) using ProNA3D. (A) Inter-chain sub-interfaces identified with d_{HA} values of 3.0 Å (left), 4.0 Å (middle), and 5.0 Å (right). Nucleotides involved in inter-chain contacts are shown in color, while non-interacting nucleotides are displayed in gray. **(B)** Secondary-structure topology plot generated by ProNA3D using default nucleotide coloring (A: red, U: blue, C: yellow, G: green). Backbone connections are indicated by black edges, and base-pairing interactions annotated according to the Leontis-Westhof classification are shown in blue. **(C)** Topology plot colored according to the inter-chain interface assignments detected by ProNA3D. **(D)** Close-up view of nucleotide A38 (blue) from chain A in the three-dimensional structure, highlighting its coordination of a Mn^{2+} ion. Adapted from Genz and Topf (2026).

3.4.3 Predicted protein-RNA complex: human Fab antibody-HIV-1 RNA

The performance of ProNA3D on computationally predicted structures was evaluated using the AlphaFold3 model of the CASP16 target M1296 retrieved from the CASP16 server (<https://predictioncenter.org>). The target structure (PDB ID: 9C75) consisted of a trimer formed by the Fab heavy chain (H), light chain (L), and HIV-1 RNA (R, Rev Response Element stem-loop II G34U mutant) (Figure 3.12A).

Application of ProNA3D ($d_{HA} = 5.0 \text{ \AA}$, $d_C = 5.0 \text{ \AA}$) to this structure revealed that the RNA interacted with both chains of the antibody. The interface between the heavy chain and the RNA comprised 23 interacting protein residues and nucleotides, whereas the light chain-RNA interface involved 5 interacting protein residues and nucleotides. A comparison of the global fold of the experimentally determined structure with the AlphaFold3 prediction (Figure 3.12B) revealed that the orientations of chains H and L were highly similar between the prediction and the resolved structure, whereas the RNA (chain R) showed noticeable differences in its global orientation. Despite this discrepancy, the interfaces identified by ProNA3D displayed a consistent interaction pattern for both the prediction and the experimentally determined structure. This observation was further supported by a high Interface Contact Similarity (ICS) score of 0.875 (Lafita et al. 2018; Kretsch et al. 2026). The ICS metric, which combines precision and recall of predicted contacts compared to the native structure, indicated a strong agreement between the predicted and reference interfacial contacts.

Analysis of the predicted interface using pLDDT (see subsection 6.3.5) showed that the interface region was predicted with high confidence (Figure 3.12B), with most residues exhibiting pLDDT values between 70 and 90 and several exceeding 90. Residue-level comparison of the experimental structure and the AlphaFold3 prediction revealed that the interface between chain L and the RNA differed by only a single residue (residue 97 in L), which did not participate in the interface in the structure but in the AlphaFold3 prediction. Similarly, the interface between chain H and RNA differed by only one residue (residue 47 in R). These results confirmed that the high-confidence interface predicted by AlphaFold3 closely aligned with the experimentally observed interactions.

To further investigate the complex predicted by AlphaFold3, ProNA3D was used to generate a topology plot of the prediction (see subsection 6.3.3), incorporating pLDDT-based node coloring and visualization of protein contacts (Figure 3.12C). The topology representation showed that both antibody chains interacted with the tip region of the RNA stem, containing a bulge in the interface. Despite being at the tip, the stem region remained largely base-paired. Similar interaction patterns have been observed in other RNA/DNA-antibody complexes (e.g., PDB ID 7TQB) and have been described in a recent preprint (Bou-Nader et al. 2025). Consistent with the interface analysis in the three-dimensional structure, pLDDT-based coloring of the topology plot showed that the protein residues interacted with regions of the RNA predicted with high confidence.

Analysis of the two-dimensional interaction network (see subsection 6.3.4) for the AlphaFold3-predicted Fab-RNA complex revealed the nucleotide A52 as a node of high-connectivity (Figure 3.12D), suggesting potential functional relevance in the context of complex formation.

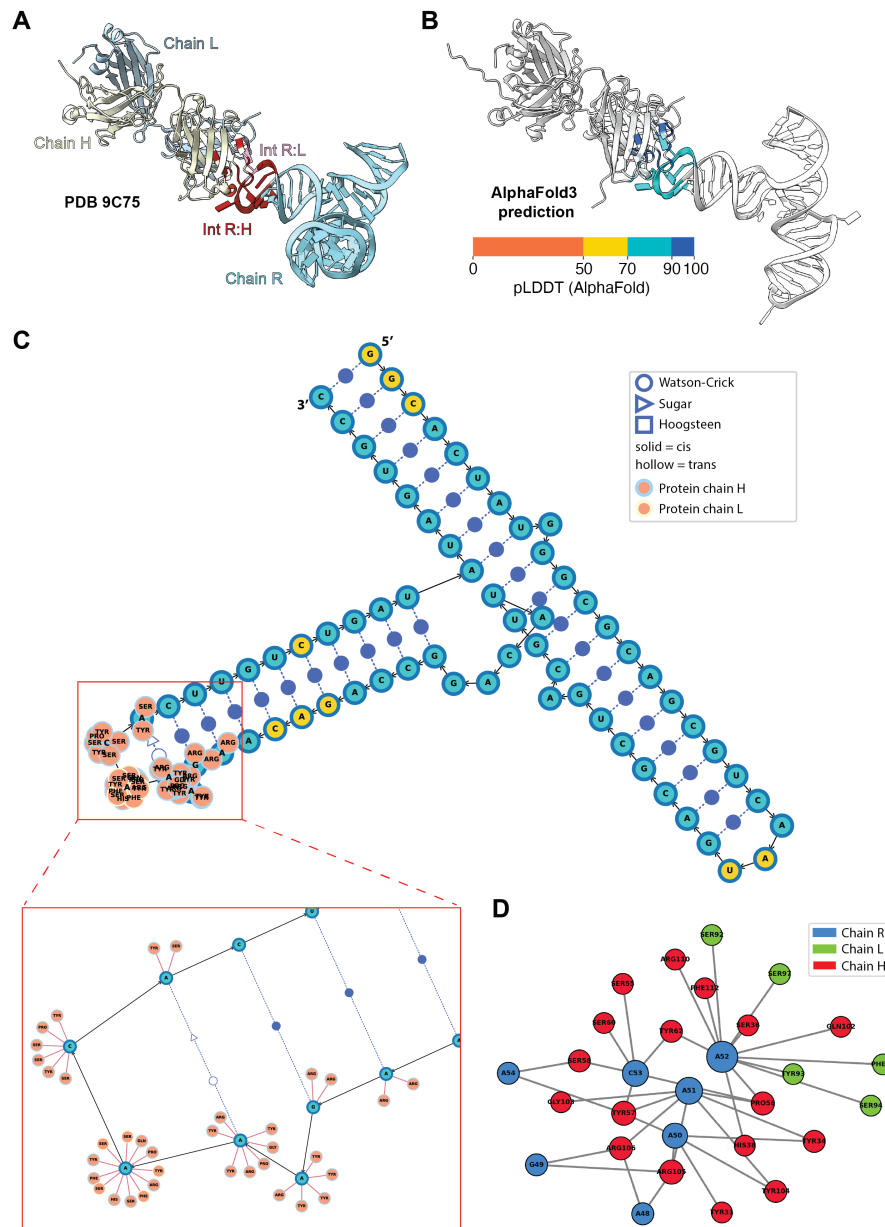


Figure 3.12: ProNA3D-based analysis of the AlphaFold3 model for the CASP16 target M1296 in comparison with its experimentally resolved structure (PDB ID: 9C75). (A) X-ray crystallography structure of the antibody-RNA complex annotated with inter-chain interfaces detected by ProNA3D. The RNA molecule (R) is shown in cyan, the Fab light chain (L) in light blue, and the Fab heavy chain (H) in beige. Interface regions are highlighted in dark red for the RNA-heavy chain interface and in pink for the RNA-light chain interface. (B) AlphaFold3-predicted model of the complex, with interface regions colored according to pLDDT confidence scores. (C) Secondary-structure topology representation generated by ProNA3D and colored by pLDDT values. Backbone connections are shown as black edges, while base-pairing interactions annotated using the Leontis-Westhof base pair classification are depicted in blue. Inset: RNA interface nucleotides and their interacting protein residues, shown in salmon. (D) Two-dimensional interface network visualization of the complex, with protein residues colored in red (chain A) and green (chain B), and RNA residues shown in blue (chain C). Adapted from Genz and Topf (2026).

3.4.4 Predicted protein-DNA complex: Cre recombinase from *Punavirus P1* bound to its target DNA

The functionalities provided by ProNA3D for the analysis of predicted protein-nucleic acid complexes and the effect of varying the pLDDT cutoff for interface residue identification were evaluated for a triple-point mutant of the Cre recombinase from *Punavirus P1* in a dimeric complex with its target DNA (PDB ID: 7RHY) (Figure 3.13A). The complex was predicted with the AlphaFold3 web server.

Using a pLDDT cutoff of 50 and default values for d_{HA} and d_C (see subsection 6.3.2), ProNA3D identified two sub-interfaces comprising 102 and 5 interaction partners (Figure 3.13B). For this cutoff, ProNA3D calculated an ipLDDT of 88.51 for sub-interface 1 and 89.62 for sub-interface 2, while the iPAE values were 2.66 Å, and 2.26 Å, respectively.

When the pLDDT cutoff was increased to 70, ProNA3D identified 95 interaction partners in sub-interface 1 and 5 in sub-interface 2 (Figure 3.13C). The resulting ipLDDT values were 89.43 for sub-interface 1 and 89.62 for sub-interface 2, and the iPAE values were 2.41 Å, and 2.26 Å.

Increasing the cutoff further to 90 reduced the interface to 16 partners in sub-interface 1 and 3 in sub-interface 2 (Figure 3.13D). For this cutoff, the ipLDDT values were 90.37 for sub-interface 1 and 90.29 for sub-interface 2, and the corresponding iPAE values were 2.08 Å, and 2.05 Å. In particular, the sub-interfaces obtained using a pLDDT cutoff of 90 exhibited high confidence, indicating reliable side-chain modeling. Such highly confident interfaces may therefore be suitable for downstream applications such as docking studies in drug discovery or targeted mutagenesis of residues involved in the interaction.

Additional global confidence metrics remained unchanged across interface definitions, as they are not recalculated by ProNA3D. These included the pTM score (0.92), the ipTM score (0.89), and the *model confidence* (0.90). The consistently high values further supported the reliability of the predicted complex.

The overall quality of the prediction was additionally confirmed by RMSD analysis using ChimeraX (Pettersen et al. 2021). Superposition of the predicted complex onto the target structure resulted in an RMSD of 1.24 Å across 249 pruned atom pairs, and an RMSD of 3.07 Å across all 301 atom pairs. Together, these results indicated that the predicted complex closely aligned with the experimentally determined structure and demonstrated that ProNA3D provides a robust framework for assessing and interpreting the quality of predicted protein-nucleic acid interfaces.

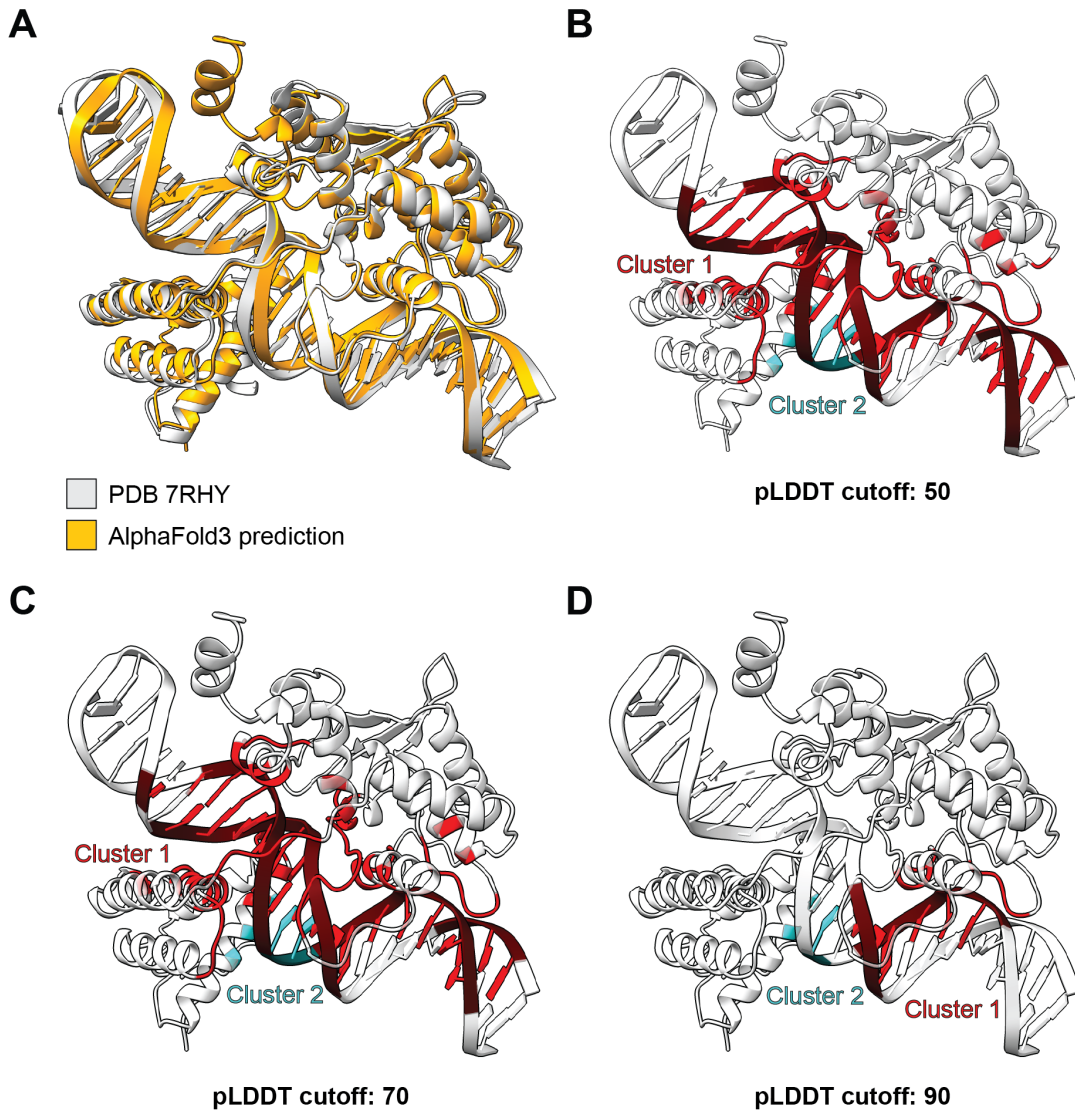


Figure 3.13: ProNA3D-based analysis of an AlphaFold3 prediction of the Cre recombinase from *Punavirus P1* in complex with its target DNA. (A) Structural alignment of the AlphaFold3 model (orange, Cre recombinase residues 1-343, DNA nucleotides 1-49) with the experimentally resolved structure (PDB ID: 7RHY, white, Cre recombinase residues 20-329, excluding unresolved regions, DNA nucleotides 1-49). (B) Inter-chain sub-interfaces identified using a pLDDT threshold of 50, highlighting sub-interface 1 (105 interacting residues and nucleotides) in red and sub-interface 2 (5 interacting residues and nucleotides) in cyan. (C) Sub-interfaces detected with a pLDDT cutoff of 70, colored as in (B), comprising 95 and 5 interacting residues and nucleotides for sub-interfaces 1 and 2, respectively. (D) Sub-interfaces identified using a pLDDT cutoff of 90, with consistent coloring as in (B), including 16 interacting residues and nucleotides in sub-interface 1 and 3 in sub-interface 2. Adapted from Genz and Topf (2026).

3.4.5 Cryo-EM application of ProNA3D

Large macromolecular assemblies that contain DNA or RNA, such as ribozymes, are frequently determined using cryo-EM due to the advantages offered by this technique (see subsection 1.2.3). To facilitate interface-focused analysis of cryo-EM complexes, ProNA3D provides an option to restrict the density map (zoning) to the identified interface sub-interfaces. This enables detailed examination of interfaces and flexible regions that often occur in low-resolution areas, particularly in composite maps and RNA-containing complexes.

This density-zoning feature (see subsection 6.3.6) was used to examine the protein-RNA complex formed by the ATP-dependent DNA/RNA helicase DHX36 from *Bos taurus* and a human G-quadruplex RNA (cMyc G-quadruplex, class I, PDB ID: 8VV2) (Figure 3.14A). The complex was deposited with a reported global resolution of 2.6 Å. Using default values for d_{HA} and d_C , ProNA3D identified two sub-interfaces within the complex (Figure 3.14B, C). Zoning the density map to sub-interface 1 revealed a high local resolution in this region of the interface, indicating reliable modeling and suggesting that even side chains can be interpreted with high confidence (Figure 3.14B, D).

In contrast, zoning the density map to sub-interface 2 showed substantially lower local resolution in this region interacting with the RNA (Figure 3.14C, D). Consequently, the modeled side chains in this area should be interpreted with caution. Closer inspection revealed that two lysine residues, Lys854 and Lys855, were only partially modeled. This was likely due to the lower local resolution in this region of the map, compared with sub-interface 1 (Figure 3.14D). However, ProNA3D classified these residues as part of the interface, thereby supporting their proposed positioning and potential involvement in RNA binding, consistent with the findings reported by Banco et al. (2025). Further analysis of the AlphaFold3-predicted complex showed that the pLDDT profile closely mirrored the local resolution pattern observed in the cryo-EM map (Figure 3.14E), with sub-interface 2 exhibiting lower predicted confidence than sub-interface 1. The combination of reduced local resolution and lower model confidence in this region may indicate increased structural flexibility.

This case illustrates how interface-based density zoning in ProNA3D provides an effective approach for analyzing protein-nucleic acid interfaces within cryo-EM density maps, which often exhibit pronounced variations in local resolution.

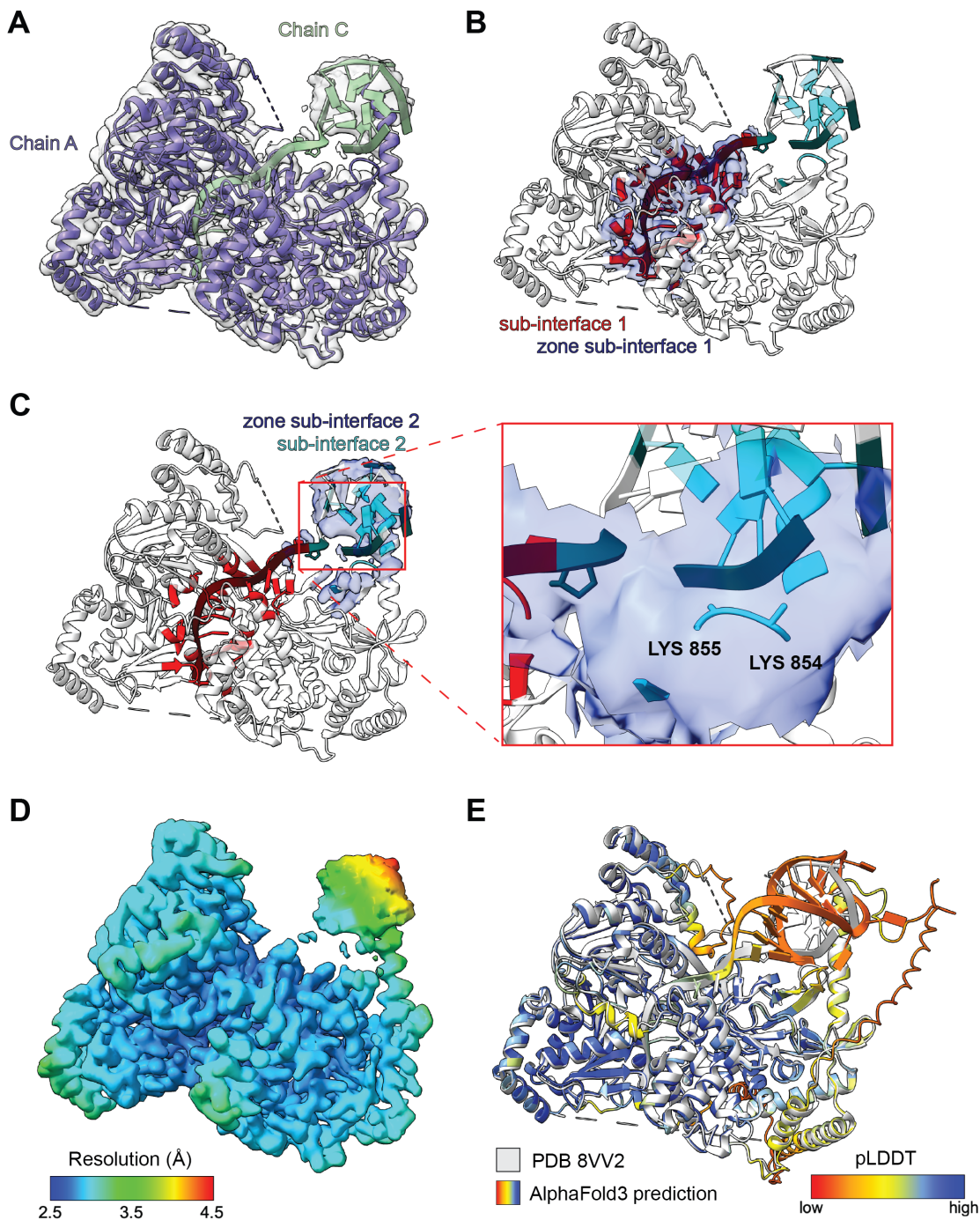


Figure 3.14: Cryo-EM-focused ProNA3D analysis of the ATP-dependent DNA/RNA helicase DHX36 in complex with G-quadruplex RNA (PDB ID: 8VV2, EMD-43547). (A) Overall view of the dimeric complex shown with the helicase in purple, RNA in green, and density map in gray. (B) Sub-interfaces identified by ProNA3D. Sub-interface 1 is shown in red, and sub-interface 2 is displayed in cyan. The density map is shown in blue around sub-interface 1. (C) Density map (blue) zoned around sub-interface 2. Inset highlights Lys854 and Lys855, which are not fully modeled due to missing density but contribute to sub-interface 2. (D) Local resolution map (Å) of the complex, with blue indicating higher resolution and red lower resolution. The map was computed using RELION-4.0.0 (Kimanus et al. 2021) with a mask applied to the modeled region. (E) Superposition of the AlphaFold3 prediction of the complex (helicase residues 1-972, RNA nucleotides 1-27), colored by pLDDT with red indicating low and blue high values on the cryo-EM structure (PDB ID: 8VV2, gray, helicase residues 15-948, RNA nucleotides 2-27, unresolved regions not modeled). Adapted from Genz and Topf (2026).

Together, these case studies demonstrated the functionalities of ProNA3D and the applicability of the plug-in to the analysis of protein-nucleic acid and nucleic acid-only complexes.

3.5 Systematic analysis of protein-nucleic acid and nucleic acid-only interfaces using ProNA3D

Previously, ProNA3D was described primarily in the context of its ChimeraX plug-in implementation. However, ProNA3D was developed not only as a visualization plug-in for the direct analysis of the results in the three-dimensional structure, but also as a command-line tool. This implementation enables rapid, large-scale analysis of macromolecular complex structures without the need for visualization.

Here, the command-line version of ProNA3D was used to analyze all high-resolution nucleic acid-containing complexes deposited in the PDB with a resolution of ≤ 3.0 Å. Using default ProNA3D settings (d_{HA} and d_C of 5.0 Å), sub-interfaces were identified across these complexes (a total of 6,047), resulting in 59,959 detected sub-interfaces. The dataset curation procedure is detailed in subsection 6.3.7.

The identified sub-interfaces were classified according to their interaction type: RNA-RNA ($n = 4,797$), RNA-DNA ($n = 450$), DNA-DNA ($n = 4,590$), protein-DNA ($n = 11,791$), and protein-RNA ($n = 23,616$). In addition, 14,681 sub-interfaces were detected between protein chains, and 34 sub-interfaces involved hybrid chains containing both protein residues and nucleotides that could not be unambiguously assigned to a single category (e.g., aminoacyl-tRNA). As ProNA3D was primarily developed for the analysis of nucleic acid-containing interfaces, these latter two interface classes were excluded, resulting in a final dataset comprising 45,244 sub-interfaces. Unless stated otherwise, the term *interfaces* in the following analyses refers to the detected sub-interfaces.

To characterize structural differences between interface classes, several properties were analyzed for all interfaces in the final dataset. These included interface size (see subsection 3.5.1), normalized B-factors (see subsection 3.5.2, see subsection 6.3.9), and maximum connectivity values (see subsection 3.5.3, see subsection 6.3.10), as well as residue- or nucleotide-specific connectivity (see subsection 3.5.4) and occurrence (see subsection 3.5.5). In addition, a curated viral subset was extracted from the full dataset and analyzed to investigate whether viral nucleic acid-containing interfaces exhibit distinct structural characteristics (see subsection 3.5.6).

Parts of the work described in this section are based on a study published as a preprint (Genz and Topf 2026), for which the manuscript is currently under review for peer-reviewed publication. The analyses of interface size and B-factor distributions represent original work conducted as part of this thesis.

3.5.1 Interface size distributions across nucleic acid-containing complexes

For RNA-RNA and DNA-DNA interfaces, interface size was defined as the total number of nucleotides participating in the interface. For mixed interfaces (protein-RNA, protein-DNA, and RNA-DNA), interface sizes were analyzed separately for each

interaction partner and reported as the number of participating protein residues (residues) or nucleotides per interaction partner type.

For RNA-RNA interfaces ($n = 4,797$), the combined interface size showed a median of 7 nucleotides (mean: 11.75, $\sigma = 17.97$) (Figure 3.15A). In contrast, DNA-DNA interfaces ($n = 4,590$) showed substantially larger interface sizes, with a median of 20 nucleotides (mean: 24.55, $\sigma = 32.97$) (Figure 3.15A). The larger median interface size observed for DNA-DNA interactions may reflect characteristic duplex packing interactions. The large standard deviations observed for both RNA-RNA and DNA-DNA interfaces indicate high variability in interface sizes within the dataset.

For RNA-DNA interfaces ($n = 450$), both interaction partners showed highly similar interface size distributions, each exhibiting a median of 9 nucleotides (RNA: mean: 10.91, $\sigma = 6.8$, DNA: mean: 10.60, $\sigma = 6.4$, Figure 3.15B). This symmetric distribution was consistent with interaction geometries dominated by base pairing and stacking interactions, which impose comparable structural constraints on both nucleic acid strands.

Furthermore, RNA and DNA interface sizes displayed a clear positive trend (Figure 3.15B), enabling deviations from the overall trend to be identified. Four pronounced outliers were observed (RNA size-DNA size: 13:6, 18:7, 19:13, 15:26) and examined in greater detail. Complexes below the general trend corresponded to large ribozymes interacting with relatively short DNA molecules. In contrast, the outlier above the trend reflected a synthetic construct, in which a relatively large DNA catalyst (27 nucleotides) component interacts with a smaller RNA (15 nucleotides).

Protein-DNA complexes ($n = 11,791$) displayed asymmetric interface size distributions for the two interaction partners (Figure 3.15C). The protein side of the identified interfaces showed a median size of 8 residues (mean: 11.11, $\sigma = 11.3$), whereas the DNA side had a median size of 4 nucleotides (mean: 4.69, $\sigma = 3.4$). This asymmetry likely reflected structural and geometric differences between proteins and nucleic acids (Arakawa et al. 2025). Individual protein residues are smaller and more conformationally flexible than nucleotides, enabling multiple residues to interact simultaneously with a single nucleotide. In contrast, nucleotides constitute larger structural units such that fewer nucleotides are typically required to form a complementary area.

Protein-RNA complexes ($n = 23,616$) showed similar nucleotide-side interface sizes but smaller protein interface contributions compared to protein-DNA complexes and exhibited substantially broader distributions (Figure 3.15D). The protein side of the interfaces had a median size of 4 residues (mean: 13.97, $\sigma = 22.8$), while the RNA side exhibited a median size of 3 nucleotides (mean: 10.73, $\sigma = 18.5$). The large standard deviations indicated a high heterogeneity in interface sizes, consistent with the diverse functional and structural roles of protein-RNA complexes (Liu et al. 2020).

Similar to RNA-DNA interfaces, protein-RNA interface sizes showed a linear trend (Figure 3.15D), enabling the identification of outliers. Due to the large number of sub-interfaces in this dataset, these outliers were analyzed in groups. The first group (10 sub-interfaces) comprised complexes with interface sizes of ≤ 50 nucleotides for the RNA side and ≥ 200 residues for the protein side. Most of the corresponding proteins were associated with CRISPR-Cas systems, including CRISPR-associated RAMP family proteins and CHAT domain proteins, interacting with short crRNA molecules. The second group (27 sub-interfaces) included complexes with interface sizes of ≥ 130 nucleotides for the RNA side and ≤ 150 residues for the protein side. The majority of these cases corresponded to 25S or 28S rRNA interacting with 60S ribosomal proteins

from 80S ribosomes across different organisms. The third group (3 sub-interfaces) consisted of complexes with RNA contributions to the interface between 60 and 90 nucleotides and the interacting protein contributing ≥ 200 residues, all representing Cas9 proteins interacting with guide RNA (gRNA).

Overall, nucleic acid-only interfaces exhibited relatively symmetric interface size distributions between interaction partners, whereas protein-nucleic acid interfaces showed more asymmetric distributions, with proteins typically contributing a larger number of interface residues than their nucleic acid partners. Additionally, protein-RNA and DNA-DNA interfaces displayed the highest variability in interface size.

The analysis of outliers further suggested that interface size patterns may, in some cases, reflect the underlying molecular class, as specific protein or RNA families tended to exhibit characteristic interface sizes.

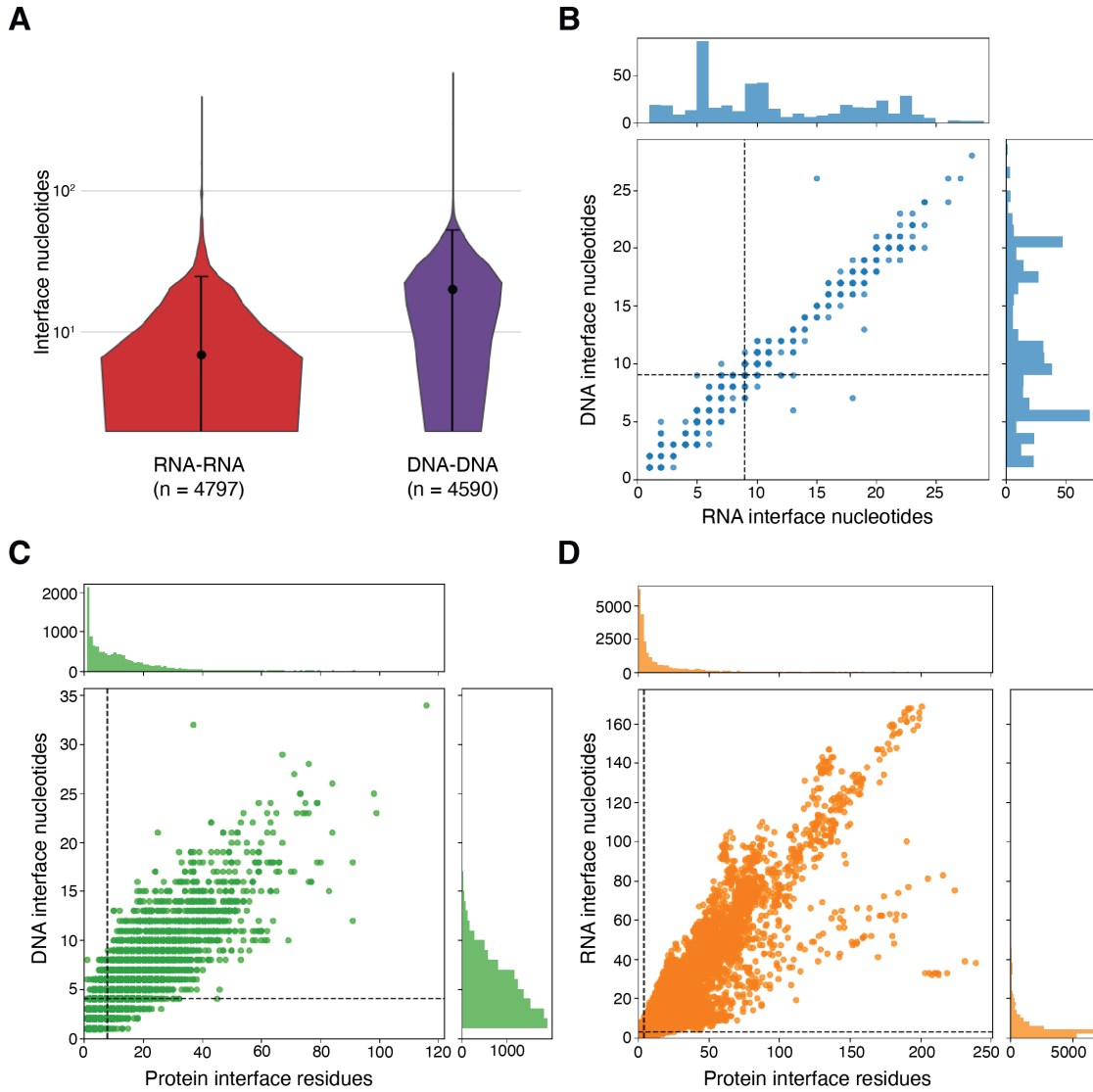


Figure 3.15: Interface size distributions across nucleic acid-containing complexes. (A) Distribution of the total number of nucleotides participating in nucleic acid-only interfaces for RNA-RNA (red, $n = 4,797$) and DNA-DNA (purple, $n = 4,590$) complexes, shown on a logarithmic scale. The median of each distribution is indicated by a black dot. (B) Number of participating residues or nucleotides per interaction partner for RNA-DNA complexes ($n = 450$). Scatter plots show the pairwise relationship between RNA and DNA interface sizes, while the marginal histograms (top and right) show the individual distributions for each chain type. Medians are indicated by black dashed lines. (C) Same representation as in (B) for protein-DNA interfaces ($n = 11,791$). (D) Same representation as in (B) for protein-RNA interfaces ($n = 23,616$).

3.5.2 Normalized B-factor analysis of interface regions

To assess the structural flexibility of interface regions in nucleic acid-containing complexes, the mean B-factor of all residues or nucleotides participating in each interface was calculated. To allow comparison across structures with different absolute B-factor ranges, B-factor values were normalized on a per-structure basis using the minimum and maximum B-factor observed within each structure (see subsection 6.3.9).

The analysis was restricted to interfaces derived from X-ray crystallographic structures at resolution better than 3.0 Å (median: 2.3 Å, mean: 2.3 Å, $\sigma = 0.5$ Å). B-factors obtained from X-ray crystallography provide a direct and physically interpretable measure of atomic displacement (Trueblood et al. 1996). In contrast, B-factors reported for cryo-EM structures are mostly model-derived and primarily reflect the local resolution (Beton et al. 2024). Consequently, restricting the analysis to X-ray structures ensured physical interpretability of the results.

The analysis revealed subtle yet consistent differences in structural flexibility between nucleic acid-only and protein-nucleic acid interactions (Figure 3.16). RNA-RNA interfaces ($n = 1,230$) exhibited a median normalized B-factor of 0.30 Å^2 (mean: 0.32 Å^2 , $\sigma = 0.16 \text{ Å}^2$), indicating moderate atomic mobility within these interfaces. DNA-DNA interfaces ($n = 4,001$) showed slightly higher values, with a median of 0.32 Å^2 (mean: 0.34 Å^2 , $\sigma = 0.15 \text{ Å}^2$).

In contrast, RNA-DNA interfaces ($n = 157$) displayed lower normalized B-factor values, with a median of 0.23 Å^2 (mean: 0.28 Å^2 , $\sigma = 0.16 \text{ Å}^2$), indicating more rigid interaction geometries. A similar trend was observed for protein-nucleic acid complexes. Protein-DNA interfaces ($n = 8,045$) exhibited a median normalized B-factor of 0.26 Å^2 (mean: 0.29 Å^2 , $\sigma = 0.14 \text{ Å}^2$), while protein-RNA interfaces ($n = 3,244$) showed nearly identical B-factor values (median: 0.26 Å^2 , mean: 0.29 Å^2 , $\sigma = 0.15 \text{ Å}^2$).

Overall, nucleic acid-only interfaces exhibited slightly higher normalized B-factors than protein-nucleic acid interfaces, indicating greater flexibility in these interfaces. In contrast, protein-nucleic acid interfaces showed lower normalized values, suggesting greater rigidity. Despite these differences, the standard deviations were similar across all interface types, indicating comparable levels of variability in local flexibility.

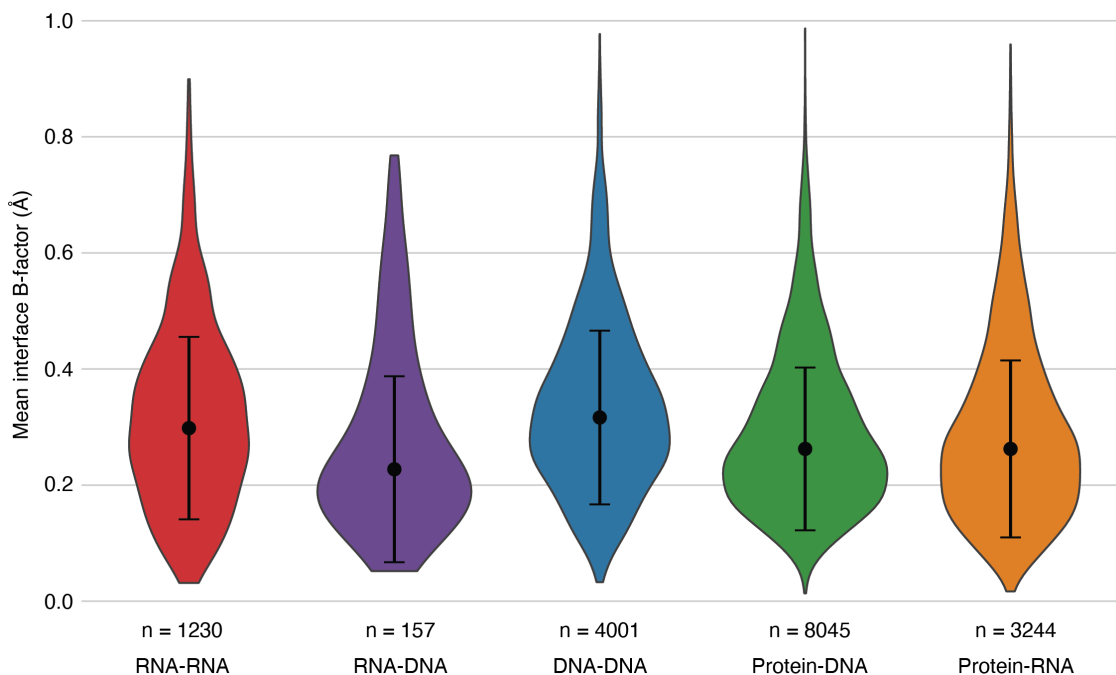


Figure 3.16: Normalized mean B-factor distributions of interface regions across interaction types in X-ray crystallographic structures. For each interface, B-factors of all participating residues or nucleotides were averaged and normalized using the minimum and maximum B-factor values within the corresponding structure. Distributions are shown separately for RNA-RNA (red), RNA-DNA (purple), DNA-DNA (blue), protein-DNA (green), and protein-RNA interfaces (orange). The median and the standard deviation for each interface type are shown.

3.5.3 Maximum connectivity distributions for different interface classes

Moreover, the maximum connectivity of interfaces in the dataset was analyzed. Connectivity of a residue or nucleotide was defined as the number of distinct residues or nucleotides on the opposing chain with which it forms at least one heavy-atom contact within d_{HA} . For each interface, connectivity was determined for every residue or nucleotide, and the highest observed value was defined as the maximum connectivity of that interface.

Nucleic acid-only interfaces displayed highly consistent maximum connectivity values across all interaction types (Figure 3.17). RNA-RNA ($n = 4,797$), RNA-DNA ($n = 450$), and DNA-DNA ($n = 4,590$) interfaces all exhibited a median maximum connectivity of 3 interaction partners per nucleotide. Mean values were similarly close (RNA-RNA: 3.22, RNA-DNA: 3.03, DNA-DNA: 2.92), indicating that maximum connectivity is relatively conserved across nucleic acid-only interfaces (Figure 3.17, Table A.12). Variability was generally low for RNA-DNA ($\sigma = 0.76$) and DNA-DNA ($\sigma = 0.88$) interfaces, whereas RNA-RNA interfaces showed a broader distribution ($\sigma = 1.83$). This likely reflected the greater structural diversity and tertiary interactions observed in RNA-only assemblies.

A local maximum at a connectivity of 1 was observed for DNA-DNA interfaces. This feature can be largely attributed to nicked DNA structures (Aymami et al. 1990), such as those generated during endonuclease cleavage. An example for this pattern

is the DNA substrate of repair enzymes (PDB ID: 1VTE, Figure 3.18A). Although the phosphodiester backbone is interrupted in these structures, base stacking interactions across the nick remain intact, resulting in non-covalent contacts corresponding to a connectivity of 1 (Figure 3.17, DNA-DNA).

In contrast, protein-nucleic acid interfaces exhibited higher maximum connectivity values and substantially broader distributions. Protein-RNA ($n = 23,616$) interfaces showed a median maximum connectivity of 4 (mean: 4.64, $\sigma = 3.32$), while protein-DNA ($n = 11,791$) interfaces displayed a median of 5 (mean: 5.52, $\sigma = 3.48$) (Figure 3.17, Table A.12). Compared to nucleic acid-only interfaces, this corresponds to approximately one and two additional interaction partners, respectively. The larger standard deviations and higher mean maximum connectivity values further indicate greater heterogeneity in interface organization, likely reflecting the wide range of protein-nucleic acid binding modes and the smaller size of protein residues compared to nucleotides (Bellone et al. 2025).

Pairwise statistical comparisons of maximum connectivity distributions confirmed that differences between nucleic acid-only and protein-nucleic acid interface connectivities were significant across interface classes (Mann-Whitney U test, see subsection 6.3.11, Table A.13). In contrast, no significant differences were observed among the nucleic acid-only interface types, highlighting the similar connectivity patterns across this group.

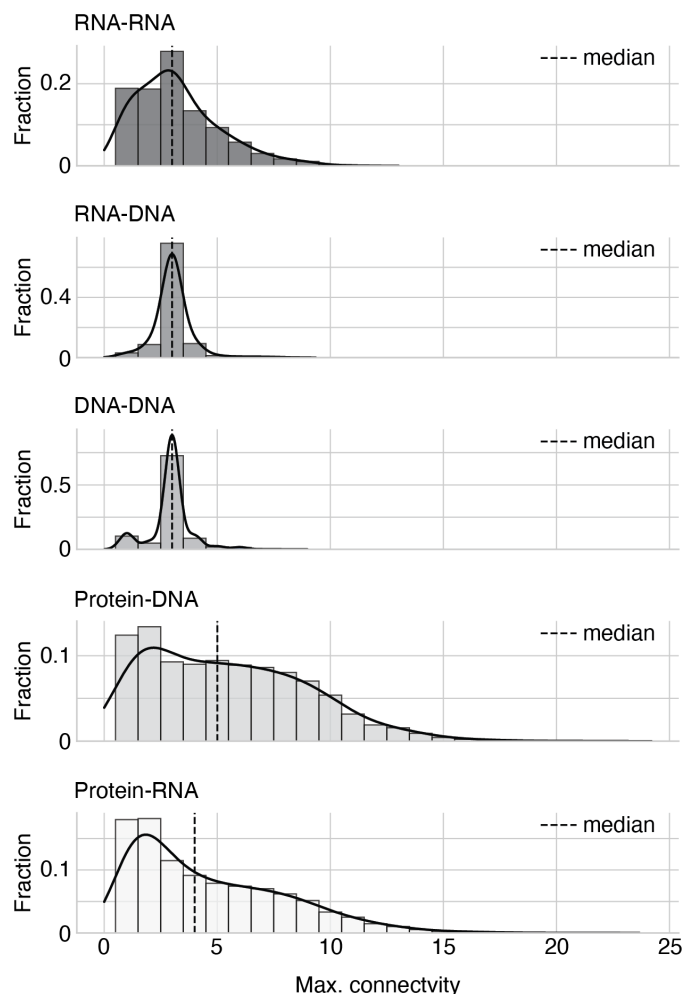


Figure 3.17: Distribution of maximum connectivity across nucleic acid-containing interface types. Connectivity was defined as the number of distinct residues or nucleotides with which a given residue or nucleotide forms interactions across an interface. For each interface, the highest observed connectivity value (maximum connectivity) was retrieved. Distributions are shown separately for RNA-RNA, RNA-DNA, DNA-DNA, protein-DNA, and protein-RNA interfaces. The median for each interface type is indicated as a dotted line. Adapted from Genz and Topf (2026).

A subset of protein-nucleic acid interfaces exhibited particularly high maximum connectivity values (Figure 3.17, Table A.12). To further characterize these cases, interfaces with a maximum connectivity ≥ 18 were analyzed in detail. Most of these outliers corresponded to enzyme-DNA complexes, particularly methyltransferases, which frequently displayed connectivity values of 20 or higher. In these systems, individual nucleotides are often flipped out of the DNA strand to enable catalytic processing, such as covalent modification or base excision (e.g., PDB ID: 2C7O) (Neely et al. 2005). In these complexes, the flipped nucleotide is typically completely surrounded by protein residues, as observed for the DNA methyltransferase from *H. haemolyticus* in complex with its target DNA (Figure 3.18B). This structural arrangement results in unusually high connectivity for a single nucleotide.

Interfaces with connectivity values between 18 and 20 were frequently associated with components of the CRISPR-Cas system. One example is the CRISPR-associated

protein Csm4 from *L. lactis* (PDB ID: 9ASH), which interacts with crRNA through a binding mode that involves base flipping of a cytosine residue (C2). This interaction pattern resulted in a connectivity of 18 (Figure 3.18C). A similar flipped conformation and high connectivity were also observed for Csm4 from *S. thermophilus* (PDB ID: 9NO4, Figure A.10), indicating that this binding mode is conserved among Csm4 proteins. These observations were consistent with previous structural studies describing nucleotide flipping at the 5' tag of the RNA, mediated by a β -hairpin motif of Csm4 (You et al. 2019).

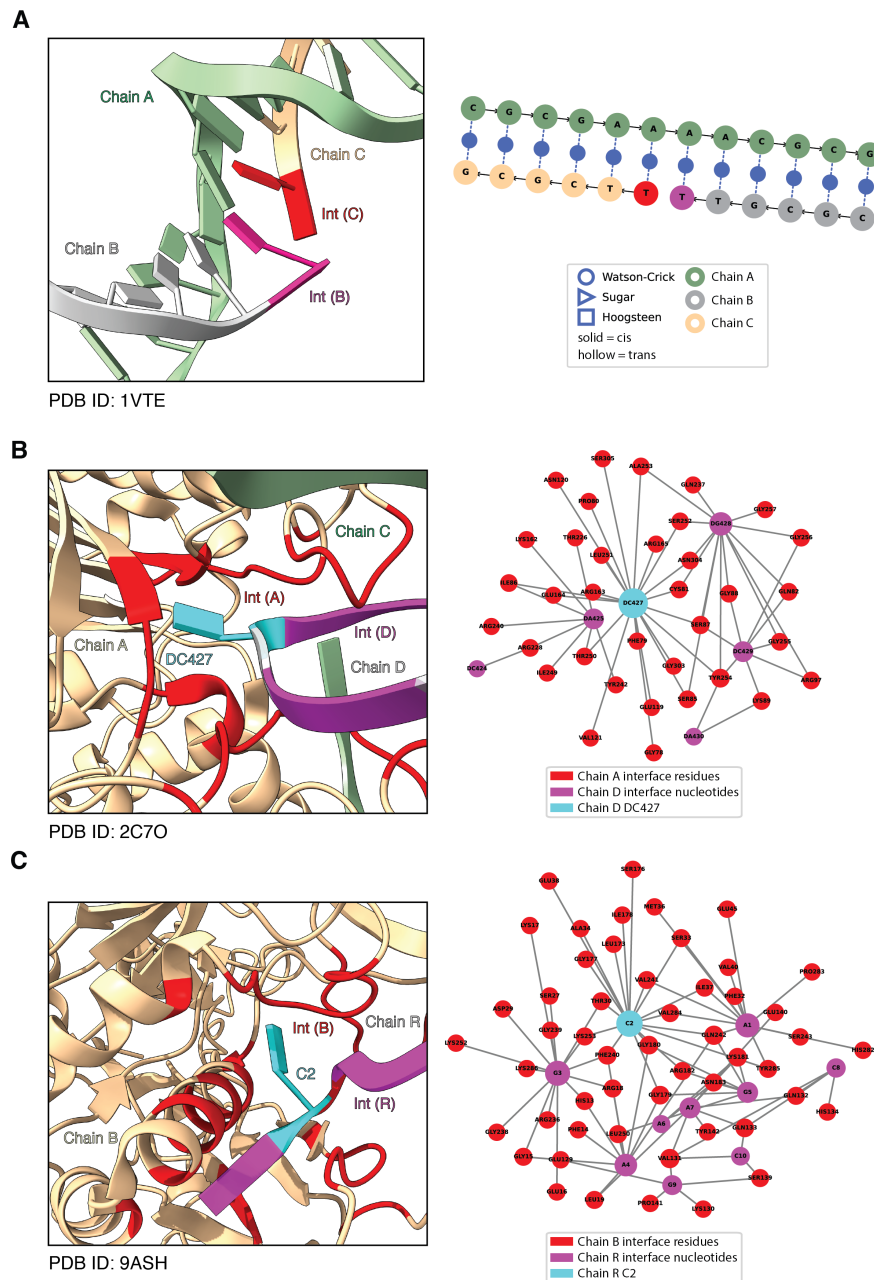


Figure 3.18: Examples illustrating low- and high-connectivity interface outliers. Shown are representative cases of nicked DNA (maximum connectivity = 1) and nucleotide base flipping associated with elevated connectivity values (≥ 18). Interface residues are colored according to interaction partner, with partner 1 in red and partner 2 in magenta. Flipped nucleotides are highlighted in cyan. For each complex, the corresponding secondary structure representation (A) and interaction networks derived from ProNA3D (B, C) are provided. (A) Nicked DNA duplex (PDB ID: 1VTE), demonstrating preserved base stacking across a discontinuity in the DNA backbone. (B) DNA methyltransferase (chain A, beige) from *H. haemolyticus* in complex with its target DNA (chain D, gray) (PDB ID: 2C7O), showing base flipping associated with dense local interactions. (C) CRISPR-associated protein Csm4 (chain B, beige) from *L. lactis* bound to crRNA (chain R, gray) (PDB ID: 9ASH), illustrating a flipped nucleotide within a highly connected interaction environment. Adapted from Genz and Topf (2026).

3.5.4 Individual residue- and nucleotide-level connectivity for different interface classes

A residue- and nucleotide-level analysis of connectivity values further highlighted the differences between nucleic acid-only and protein-nucleic acid interfaces (Figures 3.19 and 3.20, Tables A.14 to A.21).

DNA-DNA interfaces exhibited highly uniform connectivity patterns across all four nucleotides, with median values consistently at 3 and low variability ($\sigma = 0.60$ - 0.61), likely reflecting the regular and geometrically constrained nature of canonical base pairing (Figure 3.19A, Table A.14).

In comparison, RNA-RNA interfaces showed broader connectivity distributions, with median values consistently at 2 and higher standard deviations ($\sigma = 1.10$ - 1.31) (Figure 3.19B, Table A.15). Within these interfaces, the purines adenine and guanine exhibited slightly higher connectivity than the pyrimidines.

RNA-DNA interfaces showed comparable connectivity patterns for both interaction partners. Median connectivity values were distributed between 2 and 3. Specifically, adenine and guanine on the RNA side, as well as adenine and thymine on the DNA side, exhibited median connectivities of 3 (Figure 3.19C, D, Tables A.16 and A.17). The relatively low variability ($\sigma \sim 0.62$ - 0.72) indicated well-defined and structurally constrained hybrid interaction geometries.

In protein-nucleic acid complexes, substantially greater variability in connectivity was observed. For protein-DNA interfaces, median nucleotide connectivity ranged from 3 to 4 with broader distributions ($\sigma \sim 2.74$ - 2.96). Guanine and thymine exhibited slightly higher connectivity than adenine and cytosine (Figure 3.19E, Table A.18). On the protein side, connectivity varied strongly by residue type. The positively charged residues, especially arginine (median: 2) and lysine (median: 2), showed the highest connectivity, whereas hydrophobic residues such as alanine, leucine, and valine generally showed lower values (median: 1) (Figure 3.20A, Table A.19). Polar and aromatic residues also showed a median connectivity of 2, but lower mean values (1.63-1.86) compared to arginine (mean: 2.02).

A similar but more pronounced pattern was observed for protein-RNA interfaces. RNA nucleotides consistently exhibited median connectivity values of 2, accompanied by large standard deviations ($\sigma \sim 2.11$ - 2.35), which indicated substantial heterogeneity in the dataset (Figure 3.19F, Table A.20). On the protein side, arginine and lysine displayed the highest connectivity (Figure 3.20B, Table A.21). Most other residues, including polar and hydrophobic amino acids, showed a median connectivity of 2, but with broader distributions ($\sigma = 1.01$ - 1.61) than those observed at protein-DNA interfaces ($\sigma = 0.62$ - 0.80).

Overall, these residue- and nucleotide-level analyses demonstrated that nucleic acid-only interfaces are characterized by relatively uniform connectivity patterns, whereas protein-nucleic acid interfaces show greater variability. In these complexes, specific amino acid classes, particularly positively charged residues, acted as consistent interaction hotspots.

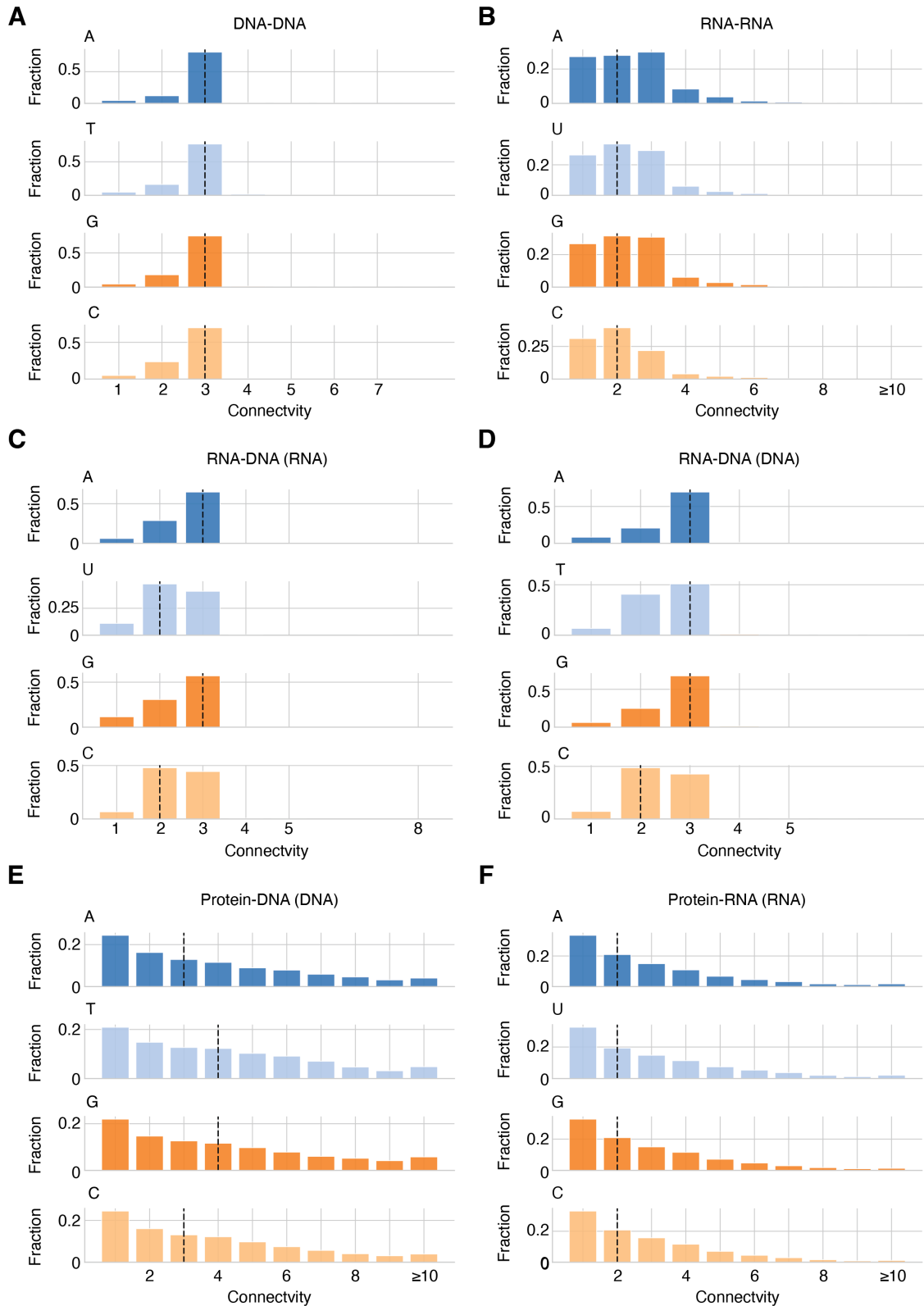


Figure 3.19: Per-nucleotide type connectivity (A/DA: blue; U/DT: lightblue; DG/G: orange; C/DC: lightorange) across nucleic acid-containing interface types. Connectivity was defined as the number of distinct residues or nucleotides contacted by a residue or nucleotide across an interface. **(A)** DNA-DNA, **(B)** RNA-RNA, **(C)** RNA-DNA (RNA), **(D)** RNA-DNA (DNA), **(E)** protein-DNA (DNA), and **(F)** protein-RNA (RNA). Median values are indicated for each interface class (dotted line). Adapted from Genz and Topf (2026).

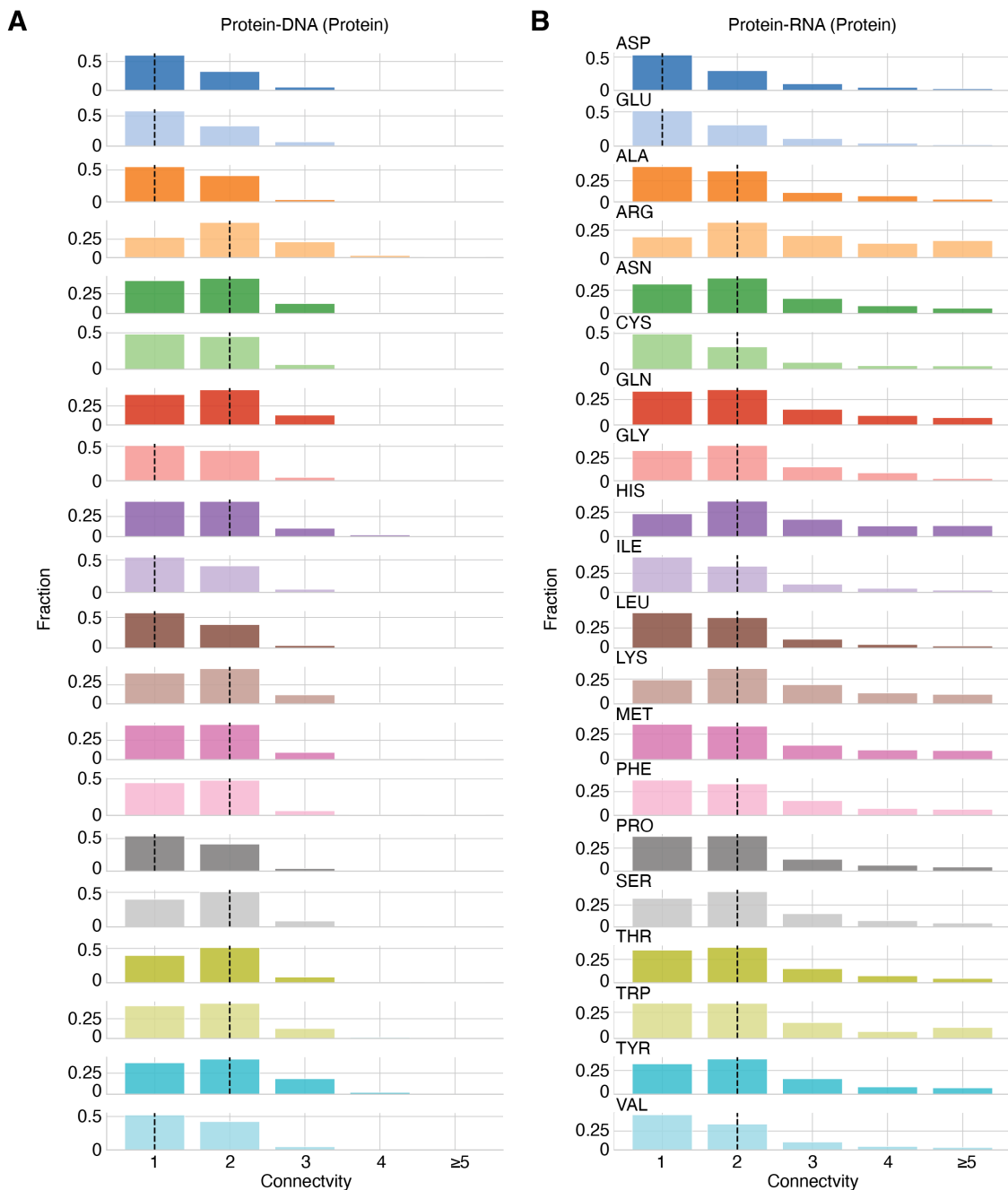


Figure 3.20: Per-residue type connectivity across nucleic acid-containing interface types. Connectivity was defined as the number of distinct residues or nucleotides contacted by a residue or nucleotide across an interface. **(A)** protein-DNA (protein) and, **(B)** protein-RNA (protein). Median values are indicated for each interface class (dotted line). Adapted from Genz and Topf (2026).

3.5.5 Occurrence distributions of residue- and nucleotide-types for different interface classes

A more detailed analysis of nucleotide and residue occurrence distributions further supported the trend observed across interface types in subsection 3.5.4. For nucleic acid-only interfaces, DNA-DNA interactions displayed a highly uniform

nucleotide composition, with all four nucleotides occurring at comparable frequencies (Figure 3.21A, Table A.14).

In RNA-RNA and in RNA-DNA interfaces, adenine and guanine occurred more frequently than cytosine and uracil/thymine, indicating a purine preference in these interface types (Figure 3.21B, C, D, Tables A.15 to A.17).

More pronounced compositional trends were observed in protein-nucleic acid complexes. Protein-DNA interfaces showed an enrichment of adenine and thymine compared to guanine and cytosine, suggesting a preference for interactions within AT-rich regions (Figure 3.21E, Table A.18). In contrast, protein-RNA interfaces showed higher frequencies of adenine and guanine (Figure 3.21F, Table A.20).

On the protein side, residue occurrence was dominated by positively charged amino acids, with arginine and lysine being the most abundant at both protein-DNA and protein-RNA interfaces (Figure 3.22A, B, Tables A.19 and A.21). Polar residues, including asparagine, glutamine, serine, and threonine, were also commonly observed, whereas hydrophobic residues occurred less frequently. This distribution was consistent with the electrostatic nature of protein-nucleic acid interactions and aligned with observations reported by Krüger et al. (2018).

Overall, the occurrence distributions supported the hypothesis that nucleic acid-only interfaces are characterized by relatively uniform interaction patterns, while protein-nucleic acid interfaces exhibit more pronounced compositional preferences, reflecting the diverse binding strategies employed by proteins.

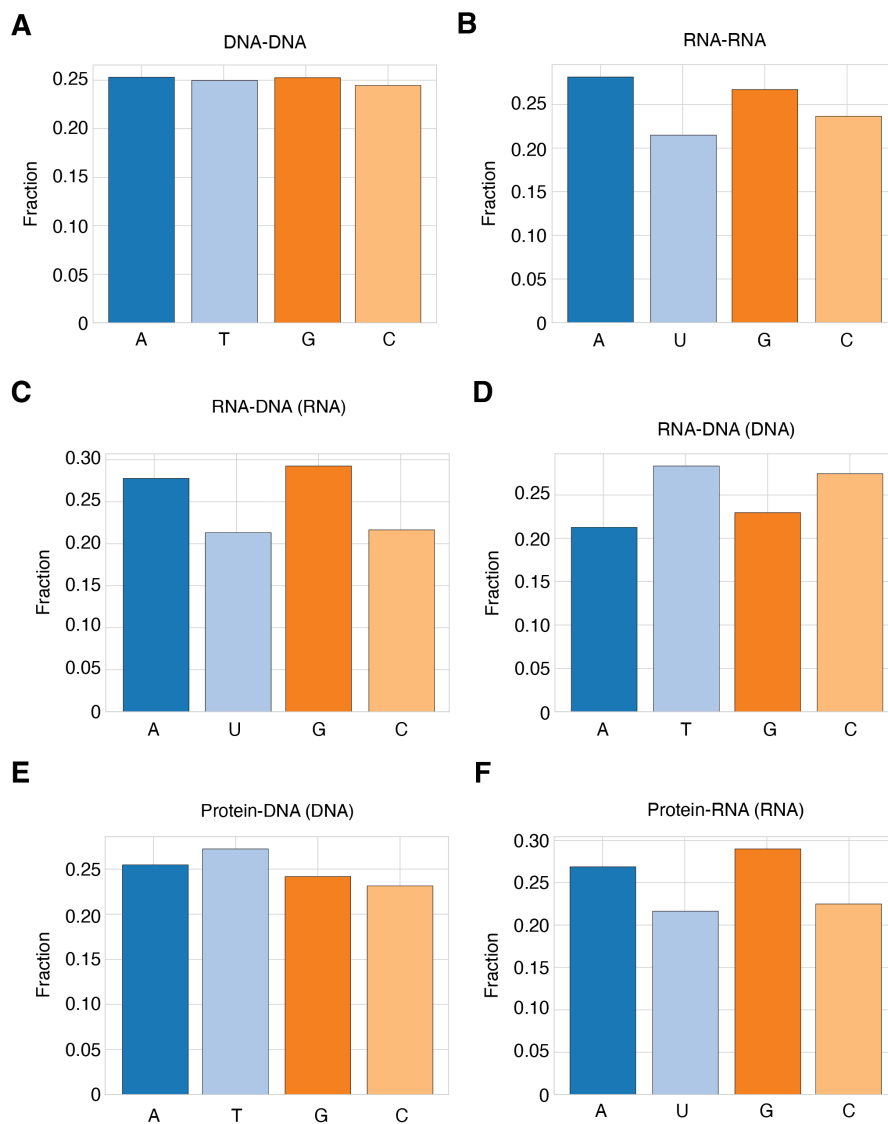


Figure 3.21: Per-nucleotide type occurrence (A: blue; U/T: lightblue; G: orange; C: lightorange) across nucleic acid-containing interface types (A) DNA-DNA, (B) RNA-RNA, (C) RNA-DNA (RNA), (D) RNA-DNA (DNA), (E) protein-DNA (DNA), and (F) protein-RNA (RNA). Adapted from Genz and Topf (2026).

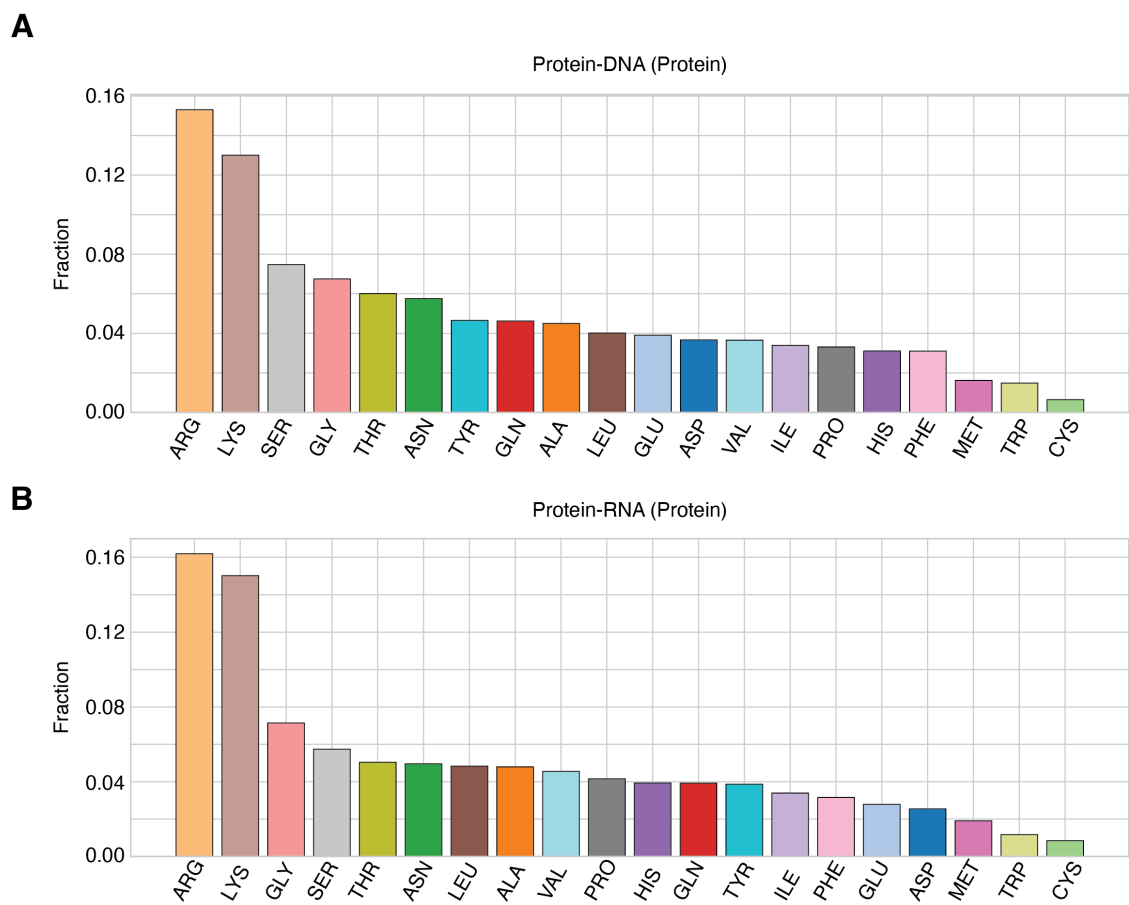


Figure 3.22: Per-residue type occurrence across nucleic acid-containing interface types. (A) protein-DNA (protein) and, (B) protein-RNA (protein). Adapted from Genz and Topf (2026).

3.5.6 Nucleic acid-containing interfaces in viral complexes

To investigate whether viral complexes exhibit distinct structural interface characteristics, a viral subset was extracted from the full dataset based on taxonomic classification (see subsection 6.3.8). The subset comprised 2,397 nucleic acid-containing sub-interfaces across all interface types (Table A.22): RNA-RNA ($n = 198$), RNA-DNA ($n = 30$), DNA-DNA ($n = 329$), protein-DNA ($n = 938$), and protein-RNA ($n = 902$). Interface size, normalized B-factors, and maximum connectivity were compared between the viral subset and the full dataset. Due to the small number of RNA-DNA interfaces in the viral dataset, results for this interface type should be interpreted with caution.

Interface sizes RNA-RNA interfaces in the viral subset exhibited a median size of 13 nucleotides (mean: 16.32, $\sigma = 14.3$), compared to a median of 7 nucleotides (mean: 11.75, $\sigma = 18.0$) in the full dataset (Figure 3.23A, Table A.22). Additionally, the distribution was narrower in the viral subset, indicating more uniform interface sizes. This may reflect the prevalence of extended, regularly base-paired double-stranded RNA regions in viruses (e.g. in dsRNA viruses). A similar trend was observed for DNA-DNA interfaces in the viral subset, with a median size of 26 nucleotides (mean: 25.78, $\sigma = 20.9$) compared to 20 nucleotides (mean: 24.55, $\sigma = 33.0$) in the full dataset (Figure 3.23A, Table A.22).

In protein-RNA complexes, the protein side of the interfaces exhibited a substantially larger median size of 7 protein residues (mean: 15.56) in the viral subset compared to 4 residues (mean: 13.97) in the full dataset, while the RNA-side contribution demonstrated only a slight increase to a median of 4 nucleotides (mean: 7.51) compared to 3 nucleotides (mean: 10.73) in the full dataset (Figure 3.23D, Table A.22). In protein-DNA complexes, the median of the DNA side of the interfaces remained unchanged (median of 4 in both datasets, mean: 5.05 vs. 4.69), while protein contributions showed a slight increase in size to a median of 9 (mean: 13.05) compared to a median of 8 (mean: 11.11) in the full dataset (Figure 3.23C, Table A.22).

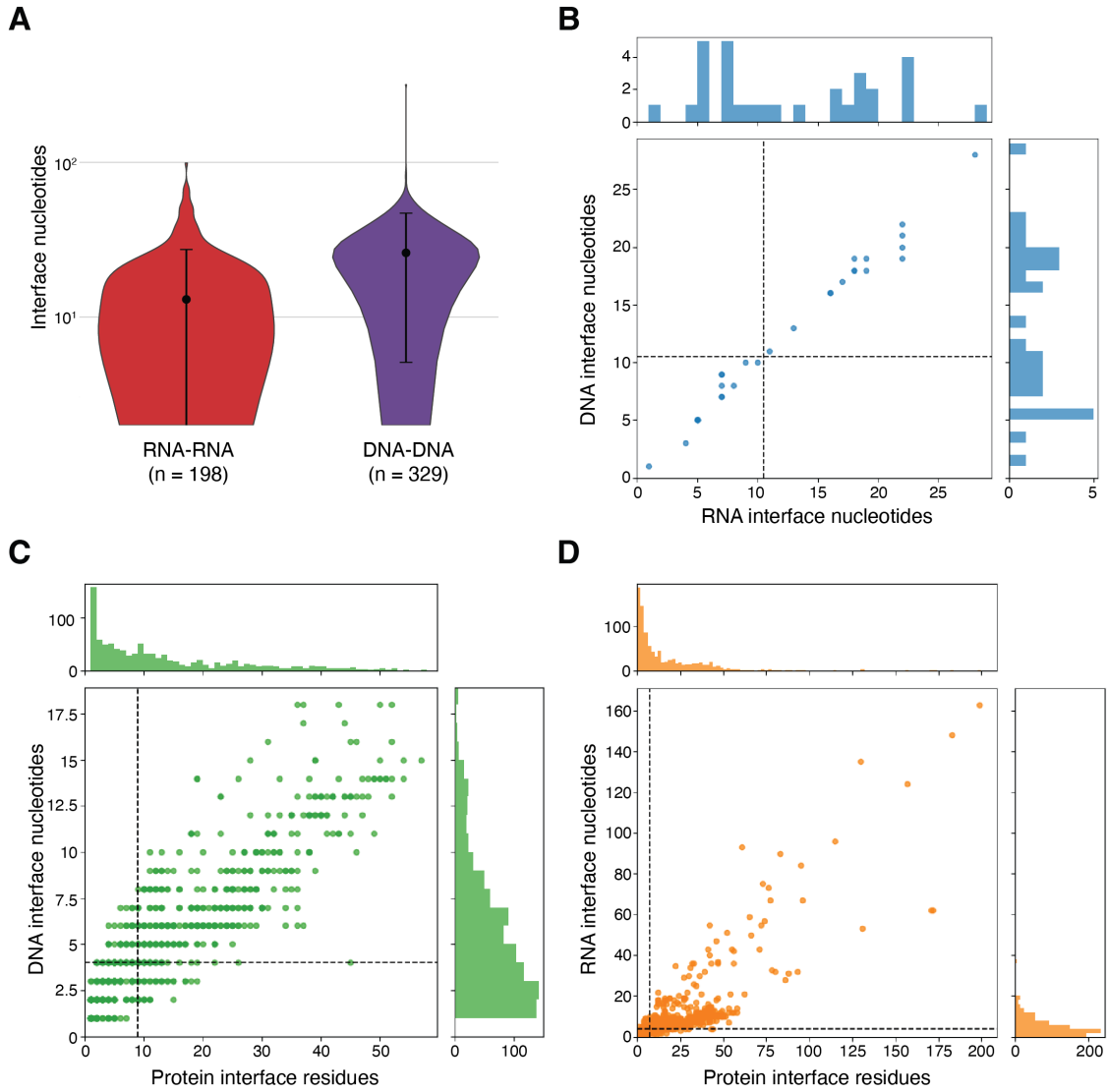


Figure 3.23: Interface size distributions across nucleic acid-containing complexes in the viral subset. (A) Distribution of the total number of nucleotides participating in nucleic acid-only interfaces for RNA-RNA (red, $n = 198$) and DNA-DNA (purple, $n = 329$) complexes, shown on a logarithmic scale. The median of each distribution is indicated by a black dot. (B) Number of participating residues or nucleotides per interaction partner for RNA-DNA complexes ($n = 30$). Scatter plots show the pairwise relationship between RNA and DNA interface sizes, while the marginal histograms (top and right) show the individual distributions for each chain type. Medians are indicated by black dashed lines. (C) Same representation as in (B) for protein-DNA interfaces ($n = 938$). (D) Same representation as in (B) for protein-RNA interfaces ($n = 902$).

Normalized B-factors. The analysis of normalized B-factors restricted to X-ray crystallographic structures (resolution median: 2.4 Å, mean: 2.4 Å, $\sigma = 0.4$ Å) revealed elevated values for viral nucleic acid-only interfaces in comparison to the full dataset (Figure 3.24, Table A.23).

RNA-RNA interfaces in the viral subset ($n = 73$) exhibited a median of 0.40 Å^2 (mean: 0.41 Å^2 , $\sigma = 0.19 \text{ Å}^2$) compared to 0.30 Å^2 (mean: 0.32 Å^2 , $\sigma = 0.16 \text{ Å}^2$) in the full dataset, indicating greater structural flexibility of these interface regions. A similar trend was observed for DNA-DNA interfaces ($n = 284$), showing a median of 0.35 Å^2 (mean: 0.38 Å^2 , $\sigma = 0.17 \text{ Å}^2$) in the viral subset compared to 0.32 Å^2 in the full dataset. A large increase of the median to 0.52 Å^2 (mean: 0.53 Å^2 , $\sigma = 0.16 \text{ Å}^2$) was also observed for RNA-DNA interfaces compared to 0.23 Å^2 in the full dataset, but with only $n = 11$ X-ray structures contributing to the viral RNA-DNA B-factor comparison, this trend should be interpreted with caution.

Interestingly, protein-RNA interfaces ($n = 260$) showed highly similar normalized B-factor distributions in the viral subset (median: 0.26 Å^2 , mean: 0.29 Å^2 , $\sigma = 0.13 \text{ Å}^2$) and the full dataset (median: 0.26 Å^2). A comparable pattern was observed for protein-DNA interfaces, with the viral subset ($n = 694$) showing a median of 0.29 Å^2 (mean: 0.32 Å^2 , $\sigma = 0.16 \text{ Å}^2$) compared to 0.26 Å^2 in the full dataset. In the full dataset, nucleic acid-only interfaces exhibited higher median B-factors than protein-nucleic acid interfaces. This trend was preserved in the viral subset, indicating that protein binding stabilizes nucleic acid interface flexibility similarly in viral and non-viral complexes.

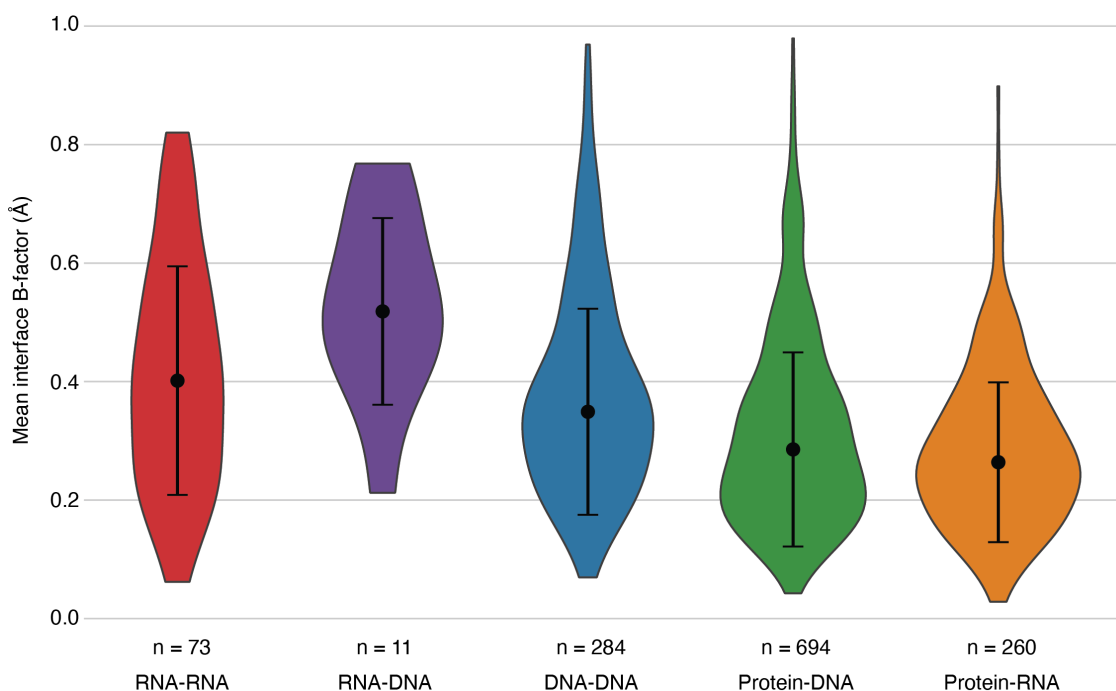


Figure 3.24: Normalized mean B-factor distributions of interface regions across interaction types in X-ray crystallographic structures in the viral dataset. For each interface, B-factors of all participating residues or nucleotides were averaged and normalized using the minimum and maximum B-factor values within the corresponding structure. Distributions are shown separately for RNA-RNA (red), RNA-DNA (purple), DNA-DNA (blue), protein-DNA (green), and protein-RNA interfaces (orange). The median and the standard deviation for each interface type are shown.

Maximum connectivity. Analysis of the maximum connectivity in the viral subset showed highly similar values for nucleic acid-only interfaces compared to the full dataset. Across RNA-RNA, DNA-DNA, and RNA-DNA interfaces, the median maximum connectivity was consistently 3 in both datasets, with comparable standard deviations (Figure 3.25, Table A.24). This observation suggests that base pairing and stacking interactions impose similar geometric constraints on nucleotide-level connectivity regardless of the origin of the complex.

In contrast, protein-nucleic acid interfaces exhibited higher maximum connectivity in the viral subset compared to the full dataset (Figure 3.25). For protein-RNA interfaces, the median maximum connectivity increased from 4 in the full dataset to 5 in the viral subset ($\sigma = 3.32$ vs. 4.13). The higher median maximum connectivity for protein-RNA interfaces, together with the interface sizes reported in section 3.5.6, suggested that proteins in viral complexes formed more extensive interfaces with RNA. Protein-DNA interfaces also showed an increase, with a median maximum connectivity of 5 in the full dataset and of 6 in the viral subset ($\sigma = 3.48$ vs. 3.46).

Overall, the viral subset was characterized by larger interface sizes, increased flexibility in nucleic acid-only interfaces, and higher maximum connectivity in protein-nucleic acid interfaces compared to the full dataset. These differences were most pronounced for RNA-containing interfaces. However, the similar normalized B-factors of protein-nucleic acid interfaces and the comparable connectivity values of nucleic acid-only interfaces in both datasets indicated that these properties are determined by fundamental physicochemical constraints rather than organism-specific differences.

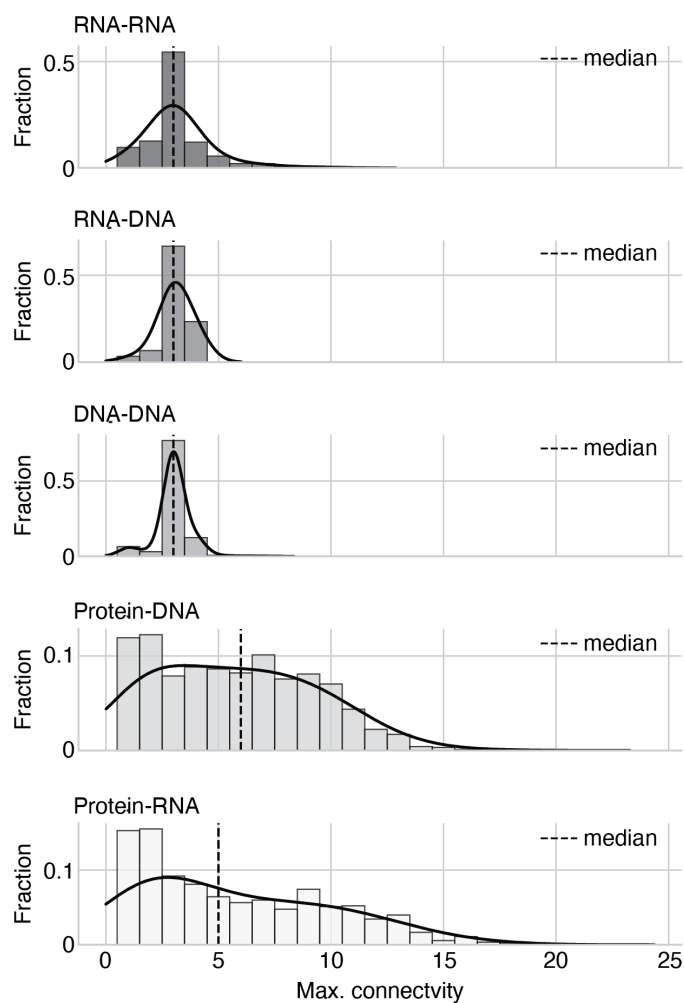


Figure 3.25: Distribution of maximum connectivity across nucleic acid-containing interface types in the viral subset. Connectivity was defined as the number of distinct residues or nucleotides with which a given residue or nucleotide forms interactions across an interface. For each interface, the highest observed connectivity value (maximum connectivity) was retrieved. Distributions are shown separately for RNA-RNA, RNA-DNA, DNA-DNA, protein-DNA, and protein-RNA interfaces. The median for each interface type is indicated as a dotted line.

4

Discussion

Contents

4.1	PICKLUSTER: spatial decomposition of protein-protein interfaces	86
4.2	Large-scale analysis of protein-protein interface properties	87
4.3	Quality assessment of predicted protein complex models	88
4.3.1	Performance hierarchy of individual scoring metrics	89
4.3.2	Assessment of prediction quality across AlphaFold3 and ColabFold	90
4.3.3	C2Qscore	91
4.4	ProNA3D: extending interface analysis to protein-nucleic acid and nucleic acid-only complexes	92
4.4.1	Detection of interaction hotspots	93
4.4.2	Confidence-guided interface analysis of AlphaFold3-predicted protein-nucleic acid complexes	94
4.4.3	Interface-based density zoning for local quality assessment in cryo-EM structures	94
4.5	Systematic analysis of nucleic acid-containing interfaces in the PDB	95
4.5.1	Interface size distributions across nucleic acid-containing complex classes	96
4.5.2	Protein binding stabilizes nucleic acid interface flexibility . .	96
4.5.3	Connectivity patterns in nucleic-acid containing complexes .	97
4.5.4	Structural features of viral nucleic acid-containing interfaces	98
4.6	Limitations	99
4.6.1	Limitations of the PICKLUSTER and ProNA3D algorithm . .	99
4.6.2	C2Qscore	100
4.6.3	Limitations of the large-scale datasets and analyses	101
4.6.4	General considerations	102

In this chapter, the results presented in chapter 3 are discussed within the broader context of structural biology and structure-based drug discovery. The first section addresses the design and application of PICKLUSTER, including the properties of protein-protein interfaces revealed by large-scale application to the Protein Data Bank

(PDB) (see section 4.1, section 4.2). Subsequently, the performance of AlphaFold-derived confidence metrics is evaluated, and C2Qscore is introduced as a combined scoring approach (see section 4.3). The third section focuses on ProNA3D and the systematic characterization of nucleic acid-containing interfaces, including a viral subset (see section 4.4, section 4.5). The chapter concludes with a discussion of methodological limitations of this work (see section 4.6).

4.1 PICKLUSTER: spatial decomposition of protein-protein interfaces

Protein-protein interfaces are often described as spatially continuous contact surfaces, although this representation fails to capture the modular organization underlying many inter-chain interactions. Characterization tools often summarize interface properties across all residue contacts into a single set of interface-level statistics, lacking the ability to identify geometrically distinct sub-regions within the interface. PICKLUSTER addresses this limitation by dividing each protein-protein interface into spatially distinct sub-interfaces via distance-based clustering of heavy-atom contacts not only in experimentally determined structures but also in AI-based complex predictions. This may be particularly relevant for structure-guided drug discovery, as hotspot residues frequently cluster within geometrically separated sub-regions of protein interfaces (Keskin et al. 2005; H. Jubb et al. 2015). Moreover, PICKLUSTER is a potentially useful tool for protein modeling challenges such as the Critical Assessment of Structure Prediction (CASP) (Kryshtafovych et al. 2019).

The definition of sub-interface-specific confidence metrics represents an additional contribution of PICKLUSTER. AlphaFold2 and AlphaFold3 provide per-residue, or per-atom predicted local distance difference test (pLDDT) scores and inter-residue predicted aligned error (PAE) values for complete structural models, but do not report these metrics at the level of individual interfaces. Although previous studies have described interface-specific variants of pLDDT and PAE (Yin et al. 2022; Teufel et al. 2023; Zhu et al. 2023), PICKLUSTER provides direct access to these scores and uniquely enables their calculation at the level of individual sub-interfaces. The interface pLDDT (ipLDDT) and interface PAE (iPAE) metrics introduced here integrate the relevant confidence information exclusively over the residues and residue pairs that constitute each sub-interface. Consequently, they reflect the confidence of the local region of interaction rather than the quality of the entire structural model. The default heavy-atom distance threshold ($d_{HA} = 5.0 \text{ \AA}$) aligned with the interface definition used by AlphaFold (Evans et al. 2022), ensuring that PICKLUSTER-derived metrics remain directly comparable to the confidence values reported by AlphaFold2 and AlphaFold3.

The *Epstein-Barr virus* (EBV) case study (see section 3.1) illustrates the importance of analyzing confidence metrics on a sub-interface level. BDRF1, BFRF1A, and BALF3 are the EBV homologues of the alphaherpesvirus terminase subunits pUL15, pUL33, and pUL28, respectively (Arakawa et al. 2025). Together, these proteins form the EBV terminase complex, which plays a central role in the viral reproductive cycle and has been identified as a promising target for antiviral drug development (Beard et al. 2002; Heming et al. 2014; Y. Yang et al. 2020; Krasnov et al. 2023). The three sub-interfaces identified between BDRF1 and BFRF1A displayed substantially different ipLDDT and

iPAE values. The sub-interface with the highest confidence (high ipLDDT and low iPAE) corresponded to a region containing potential salt-bridge and hydrogen-bond interactions, suggesting that the increased confidence may reflect well-defined structural contacts. Without division into sub-interfaces, an interface-specific confidence metric would average these contributions, preventing identification of differences that may be important for prioritizing interface regions for downstream analyses. These analyses are therefore particularly relevant for complexes of therapeutic interest, such as the BDRF1-BFRF1A interaction.

Beyond interface scoring, PICKLUSTER enables mapping of sub-interface residues onto the protein sequence, allowing for comparison of contact patterns across structurally diverse protein complexes. Proteins that do not share overall fold similarity may yet exhibit similar sub-interface patterns (Gao and Skolnick 2010). Therefore, this capability may facilitate the identification of protein families that use similar interaction strategies and allow for the investigation of related but experimentally uncharacterized complexes.

PICKLUSTER also enables the identification of hotspot residues, here defined as residues participating in multiple specific interaction types or contributing to more than one interface within a complex of higher stoichiometry. An example is Arg67 in the BDRF1-BFRF1A complex from EBV. Such residues are of particular interest for structure-guided drug discovery, as they can be critical for stabilizing protein-protein interactions and can therefore strongly influence binding affinity. Mutational analysis through alanine-scanning experiments could be used to assess the contribution of these residues to interface stability and to evaluate their functional importance. In addition, the detected interfaces could be used as targets for ligand-docking approaches, as key residues within the interface can be identified.

Representative examples of PICKLUSTER applications described in recent studies (Belios et al. 2025; Dan-Dukor et al. 2025; Mohan et al. 2025) and reported download statistics demonstrate that PICKLUSTER is an accessible tool that is already used in structural studies of protein complexes.

4.2 Large-scale analysis of protein-protein interface properties

The large-scale application of PICKLUSTER to 19,821 PDB entries provided a systematic characterization of protein-protein interface organization across the experimentally resolved structural proteome. The resulting distributions may serve as a reference for the interpretation of individual interfaces and for understanding general features of protein-protein interactions.

The observed sub-interface size distribution was right-skewed, with a median of 16 residues and a 90th percentile of 60 residues (see section 3.2). This pattern, together with the broad range of interface sizes, was consistent with previously reported distributions of protein-protein interfaces (Jones and Thornton 1996; C. Yan et al. 2008). While many studies describe interface size in terms of the buried surface area (BSA), typically reporting values in the range of 1200-2000 Å² (Lo Conte et al. 1999; Chakrabarti and Janin 2002; C. Yan et al. 2008), BSA values can be related to residue counts. Based on an estimated average contribution of 36 Å² per protein residue (Chakrabarti and Janin

2002; Janin et al. 2008), this corresponds to approximately 33-56 residues per interface. (Chakrabarti and Janin 2002) reported an average interface size of 57 ± 22 residues per protein-protein interface. In comparison, the mean size observed using PICKLUSTER was 29 residues per sub-interface, which was near the lower bound of the reported range. This difference likely reflected that approximately 40% of the analyzed protein complexes contained two or more sub-interfaces. This hypothesis was confirmed by calculating the total interface size without decomposition into sub-interfaces, yielding a mean of 48.84 residues (median: 38, $\sigma = 43.73$, $n = 29,288$ interfaces), consistent with values reported in the literature.

The discrepancy between the median (16 residues) and the mean (29 residues) indicated that the distribution was influenced by a limited number of large sub-interfaces, whereas the majority were relatively small. From a drug discovery perspective, this predominance of small sub-interfaces is particularly relevant, as such interfaces are more likely to form well-defined pockets or grooves that can be targeted by small molecules or constrained peptides (Wells and McClendon 2007; Koes and Camacho 2012). In contrast, larger interaction sites are generally more difficult to disrupt through non-covalent mechanisms (H. Jubb et al. 2015).

Approximately 60% of protein-protein interfaces in the dataset consisted of a single continuous interface, indicating that, for the majority of inter-chain contacts in the PDB, the conventional representation as a coherent interface was sufficient. However, the results showed that ~40% of interfaces comprised two or more spatially distinct sub-interfaces. These differed significantly in residue composition, BSA, and physicochemical properties, and treating them as a single entity would fail to capture this heterogeneity.

In a small number of cases, interfaces were divided into up to 18 sub-interfaces between a single pair of chains, as observed for a synthetic three-helix bundle construct (PDB ID: 4TQL). Such extreme cases likely reflected designed or repetitive structural contacts rather than typical physiological interactions. Notably, the ability of PICKLUSTER to process these cases demonstrated that the clustering algorithm is robust to structural extremes. For complexes of high stoichiometry, the detection of 4,844 sub-interfaces within a single assembly (PDB ID: 6KGX) further highlighted the scalability of the command-line implementation. This capability is essential for the analysis of large, highly stoichiometric complexes and supports the application of PICKLUSTER in systematic structural proteomics.

Overall, the large-scale analysis provided a robust reference for interface size and spatial organization, which can be useful for comparative studies.

4.3 Quality assessment of predicted protein complex models

Accurate quality assessment of predicted protein complex models is crucial for their reliable interpretation in structure-based drug discovery and other downstream analyses. ColabFold (AlphaFold2) and AlphaFold3 each provide a range of confidence metrics as standard outputs, but their comparative discriminatory power with respect to interface quality has not been systematically characterized across different prediction methods. Importantly, these scores also lack binary classification cutoffs. A comprehensive

evaluation of widely used confidence metrics (ipLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF) was therefore conducted on a curated benchmark of 223 high-resolution heterodimeric complexes from the PDB. Moreover, prediction quality was evaluated for the three prediction methods ColabFold with templates (CF-T), ColabFold without templates (CF-F), and AlphaFold3 (AF3). DockQ was applied as the reference measure of interface quality in accordance with CAPRI classification criteria (Basu and Wallner 2016). The benchmark set was composed of structures deposited after the AlphaFold3 training data cutoff date. However, entries with structural homologs among pre-cutoff PDB entries were retained to reflect realistic evaluation conditions, as predictions of complexes with related structures in the training data represent a common use case. The following discussion is based on Genz et al. (2025a).

4.3.1 Performance hierarchy of individual scoring metrics

The systematic evaluation of the seven confidence metrics against DockQ (Basu and Wallner 2016) revealed consistent differences in performance across CF-T, CF-F, and AF3 predictions (see subsection 3.3.3, subsection 3.3.4). Among these metrics, ipTM and *model confidence* showed the strongest agreement with DockQ, as they exhibited the highest Pearson correlation coefficient (PCC) and area under the curve (AUC) values in receiver operating characteristic (ROC) analyses across all datasets. Both metrics are derived from the PAE matrix (Evans et al. 2022; Dunbrack 2025). Specifically, ipTM is computed from inter-chain PAE values using weights that emphasize residue pairs with lower predicted errors, whereas *model confidence* is defined as a combination of ipTM and pTM. The consistently strong performance of these metrics suggests that the PAE matrix effectively captures inter-chain interface accuracy. This is further supported by the high correlation of iPAE with DockQ. Despite lacking explicit weighting, it averages PAE values across interface residue pairs and still provides strong discriminatory power, consistently ranking third among the scoring metrics. Consistent with these results, strong correlations were observed between PAE and pTM, as well as between ipTM and iPAE, reflecting their shared dependence on the PAE matrix.

A limitation of the standard ipTM score is its dependence on total chain length (Bret et al. 2024; Dunbrack 2025). In sequences with substantial disorder or peripheral domains, this can reduce the contribution of interface residues to the score and lead to artificially low ipTM values. The recently developed ipSAE addresses this by including only residue pairs with favorable inter-chain PAE values in the calculation and by adjusting the length normalization and scoring scheme (Dunbrack 2025). On the benchmark dataset used in this thesis, ipSAE showed slightly higher correlation with DockQ for CF-T and CF-F predictions and comparable performance for AF3 (Dunbrack 2025). Given the high-resolution heterodimeric structures in this dataset, where extensive disorder is less frequent, this limited improvement is consistent with ipSAE being most beneficial for full-length sequences with substantial disorder.

Across all three datasets, interface-specific confidence metrics consistently outperformed their global counterparts. In particular, iPAE showed stronger correlation with DockQ than global PAE, and ipLDDT outperformed pLDDT, in agreement with Yin et al. (2022). This difference can be explained by both the design of the metrics and the definition of the reference measure, DockQ. While the global pLDDT primarily reflects the folding accuracy of individual chains, the global PAE incorporates inter-chain relationships but averages error estimates over the entire complex. Consequently,

regions of low confidence outside the interface contribute to the overall score, reducing its sensitivity to interface-specific accuracy. In contrast, interface-specific metrics restrict the analysis to residues involved in direct inter-chain contacts. Furthermore, this trend was consistent with the use of DockQ as the reference metric, as two of its three components emphasize interface quality rather than global structural accuracy. Accordingly, metrics that focus on interface regions showed stronger agreement with DockQ than global measures.

Among the interface-specific scores, pDockQ2 consistently showed the lowest PCC across all three datasets, which may be explainable by the pronounced accumulation of near-zero values that exhibit low correspondence with the observed DockQ scores. This affects its ability to discriminate between correct and incorrect models, reflected in lower AUC values compared to the other individual scores and the unusually low classification thresholds derived from ROC analysis (CF-T: 0.15, CF-F and AF3: 0.05, Figure 3.6, subsection 3.3.4). This trend may be explained by the design of pDockQ2, which combines interface-specific PAE and pLDDT via a parameterized sigmoid function. This transformation compresses score values towards both the lower and upper regions of the sigmoid, reducing score variance and thereby influencing correlations with other metrics. As the metric was trained on a dataset that includes not only dimeric but also higher-order complexes, it may be optimized for systems of higher stoichiometry, which could further explain its reduced performance on dimeric interfaces. Representative structural cases showed that ipTM correctly identified high-confidence interfaces for models to which pDockQ2 simultaneously assigned very low scores, indicating that application of pDockQ2 may lead to an underestimation of interface quality.

The thresholds determined in this study for all individual scoring metrics provided practical guidance for assessing the reliability of protein complex models, taking into account the chosen prediction method and settings. This enabled the distinction between correct and incorrect models, even for predictions with ipTM values within the gray zone (Gomez and Kovalevskiy 2025a).

The sensitivity analysis of ipLDDT and iPAE to variations in the heavy-atom distance threshold ($d_{HA} = 5.0\text{--}9.0\text{ \AA}$) revealed no substantial differences in the resulting values across this range. This robustness supported the choice of $d_{HA} = 5.0\text{ \AA}$ as the default (see section 4.1) and indicated that the interface-specific confidence metrics were not critically dependent on the selected cutoff distance. Furthermore, the alignment of the d_{HA} default with the cutoff used internally by AlphaFold2 for the definition of the interface (Evans et al. 2022) ensured that PICKLUSTER-derived metrics remained directly comparable to the confidence values reported by the prediction methods.

4.3.2 Assessment of prediction quality across AlphaFold3 and ColabFold

Among the three evaluated prediction methods, AlphaFold3 generated the highest proportion of high-quality models and the lowest proportion of incorrect models (see subsection 3.3.1). CF-T ranked second in both categories, while CF-F exhibited the lowest proportion of high-quality models and the highest rate of incorrect predictions. These results indicated AlphaFold3 as the most reliable prediction method for heterodimeric complexes, consistent with the reported architectural improvements over the AlphaFold2 framework (Abramson et al. 2024). In addition, AlphaFold3 exhibited the most consistent

prediction behavior, with the largest fraction of targets for which all five generated models were uniformly classified as incorrect. This suggested that prediction failures (incorrect models) in AlphaFold3 were more systematic and were therefore more likely to be associated with consistently low confidence scores. In contrast, CF-T and CF-F displayed greater variability in model quality across the five predictions generated per target.

Despite yielding the lowest proportion of high-quality models, CF-F exhibited the highest mean PCC between individual confidence metrics and DockQ (see subsection 3.3.4). In contrast, AlphaFold3 showed the weakest mean correlations. The differences between both ColabFold datasets and AlphaFold3 were statistically significant, whereas no significant difference was observed between CF-T and CF-F. Although the scoring metric cutoffs derived in this study were also efficient for AlphaFold3 predictions, their performance was reduced compared to CF-T and CF-F, as reflected by lower AUC and accuracy values (Figure 3.6B, Table A.2). The observed discrepancies may be attributable to differences in the model architecture and confidence estimation between AlphaFold3 and AlphaFold2-Multimer (Abramson et al. 2024). Additionally, prediction performance is known to be dependent on the depth and diversity of the multiple sequence alignment (MSA). ColabFold uses MMseqs2 (Steinegger and Söding 2017), whereas AlphaFold3 employs a distinct MSA generation pipeline, which may lead to differences in alignment quality and, consequently, prediction performance.

The analysis of agreement in prediction quality across the three prediction methods showed that a high fraction of correct models was consistently identified by all approaches, whereas agreement for incorrect predictions was significantly lower. Examination of the acceptable-quality category further revealed that agreement between methods decreased with declining prediction quality. This pattern suggests that, particularly for models near the threshold between correct and incorrect classifications, combining results from multiple prediction methods may increase the likelihood of identifying a reliable structure. Notably, AlphaFold3 yielded the largest proportion of uniquely correct and high-quality predictions, further supporting its superior prediction performance among the three prediction methods.

Investigation of structural properties associated with consistent prediction failure or success revealed that the interface size and flexibility were important determinants of model quality (see subsection 3.3.5). Predictions classified as incorrect were characterized by smaller interfaces and elevated interface B-factors. The relationship between interface size and prediction accuracy was consistent with previous findings reported by Bryant et al. (2022) and Yin et al. (2022). In AlphaFold-based models, inter-chain interactions are inferred from sequence-derived signals. Larger interfaces are therefore more likely to be predicted accurately, as they provide a greater number of inter-chain contacts, increasing the amount of information available to guide correct chain positioning (Bryant et al. 2022).

4.3.3 C2Qscore

To integrate complementary information encoded by individual scoring metrics and improve robustness, ipLDDT, iPAE, pTM, ipTM, and VoroIF were combined into a single score, C2Qscore, using a linear regression model with weights optimized against DockQ on the benchmark dataset.

C2Qscore improved model quality evaluation, particularly for CF-F and AF3, while achieving performance comparable to the unweighted combination of the scoring metrics for CF-T on the X-ray test set. The advantage of C2Qscore arose from its integration of multiple complementary confidence measures, which reduced dependence on a single metric and enhanced robustness, especially in cases where individual scores provide ambiguous signals. This observation was consistent with previous studies demonstrating that combined scoring approaches, including DockQ itself (Basu and Wallner 2016), consensus-based methods in CASP (Kryshtafovych et al. 2014), and multi-metric scoring approaches for protein-protein docking (Moal et al. 2013), generally yield more robust assessments than individual metrics.

Although individual scores such as ipTM or iPAE may outperform C2Qscore in specific cases on the X-ray crystallography test set, the combined approach provided more consistent performance overall, particularly with respect to precision and specificity. However, given the limited size of the test set, these results should be interpreted with caution.

Evaluation on two independent test sets revealed an important limitation of C2Qscore and the individual scores for more complex biological contexts. Whereas performance on the X-ray crystallography test set was generally consistent with benchmark results, a decrease in classification accuracy was observed for dimers extracted from larger cryo-EM assemblies, with an elevated false-negative rate across all three prediction methods.

This degradation was attributable to the existence of multiple structurally possible dimeric configurations within higher-order assemblies. When two chains forming a dimer are embedded in a larger complex, homologous chains within the assembly may give rise to alternative configurations. As structure prediction tools receive only sequence as input, a generated model may accurately reproduce an alternative configuration that does not match the designated target used for DockQ computation. The resulting low DockQ score does not reflect poor prediction quality but rather a mismatch between the modeled and the annotated configuration. In such cases, AlphaFold2 and AlphaFold3 can still generate highly accurate models, yet assigning confidence scores ranging from medium to high. This pattern suggests that when multiple configurations of a dimer are possible, prediction methods struggle to distinguish probability from certainty, resulting in reduced confidence independent of the actual structural accuracy and contributing to the increased false-negative rate observed in the large assembly test set.

4.4 ProNA3D: extending interface analysis to protein-nucleic acid and nucleic acid-only complexes

As outlined in subsection 1.5.2, existing computational analysis tools for nucleic acid-containing complexes are typically specialized for either RNA or DNA, and tend to focus on either secondary structure annotation or interface characterization rather than providing both in an integrated framework (Figure 1.11). Moreover, these tools are mainly designed for experimentally determined structures, focus on a single type of biomolecule, and do not offer unified workflows for mixed complexes within a visualization environment.

ProNA3D addresses these limitations by providing a unified analysis platform implemented both as a UCSF ChimeraX plug-in and as a command-line tool, supporting both experimentally resolved and computationally predicted models across the full range of protein-nucleic acid and nucleic acid-only complex types. It integrates three-dimensional sub-interface identification, two-dimensional interface network graphs, secondary-structure topology plots, interface-specific confidence metrics for AlphaFold3-generated models, as well as density zoning for interface-focused local quality assessment in cryo-EM maps. The following discussion section is based on a study published as a preprint (Genz and Topf 2026), for which the manuscript is currently under review for peer-reviewed publication.

4.4.1 Detection of interaction hotspots

A central challenge in the structural analysis of nucleic acid-containing complexes is the identification of residues or nucleotides that are critical for complex formation and stability. Interface contacts do not contribute equally, as some positions mediate multiple simultaneous interactions and function as structural hubs whose disruption is likely to affect the stability of the entire assembly. Identifying such interaction hotspots is important for functional interpretation and provides a rational basis for prioritizing residues in mutagenesis experiments or for targeting interfaces in structure-guided drug discovery. ProNA3D addresses this challenge by quantifying the number of inter-chain contacts per residue or nucleotide across all detected interfaces and identifying highly connected positions within the interaction network.

The topology analysis of the dimeric hammerhead ribozyme (PDB ID: 2OEU) identified nucleotide A38 in chain A as a highly connected interaction hub (see subsection 3.4.2). This is consistent with its role in coordinating an Mn^{2+} ion, which is critical for the catalytic activity of the complex (Martick et al. 2008). The structural importance of A38 was therefore derived directly from connectivity analysis alone, without prior knowledge of the metal-binding site.

Whereas ipLDDT and iPAE values for protein-protein interfaces showed no substantial sensitivity to d_{HA} variations in the 5-9 Å range (see section 4.1), the number of sub-interfaces detected in this RNA-only complex changed substantially across the narrower range of 3-5 Å. Thresholds above 5 Å were not evaluated, as a d_{HA} of 5 Å already resulted in detection of a single continuous interface (see subsection 3.4.2), limiting the interpretability of higher thresholds for sub-interface analysis. The greater sensitivity observed for the narrower range likely reflects that many specific non-covalent interactions in macromolecular interfaces occur within this short-distance range. For example, strong hydrogen bonds between heavy-atom donor and acceptor pairs typically occur at distances below approximately 3.5 Å. A cutoff of 3.0 Å may therefore exclude relevant hydrogen bonds (Hubbard and Haider 2010). Similarly, salt bridges between Glu and Asp can span distances of about 2.5 to 4.0 Å (Ban et al. 2019). Accordingly, small d_{HA} values can substantially alter the detected interface.

The importance of network-based analysis was also demonstrated for a predicted human Fab-antibody-HIV-1 RNA complex (PDB ID: 9C75), in which nucleotide A52 was identified as a highly connected interaction node within the RNA chain (see subsection 3.4.3). The identification of A52 as a potential hotspot in a predicted complex was particularly relevant from a therapeutic perspective, as antibodies

recognizing specific RNA structural elements may serve as targets for antiviral intervention (J.-D. Ye et al. 2008).

Taken together, these examples showed that connectivity-based hotspot detection can be applied to both experimentally determined and computationally modeled structures, and across structurally diverse complexes, thereby offering a basis for downstream functional and therapeutic studies.

4.4.2 Confidence-guided interface analysis of AlphaFold3-predicted protein-nucleic acid complexes

Recent progress in AI-based approaches for modeling of protein-nucleic acid complexes has expanded the availability of structural data to a substantially broader range of assemblies, including those for which experimental characterization remains technically challenging or infeasible (Abramson et al. 2024). This development introduces new requirements for structural analysis methods. Tools developed for experimentally resolved structures typically lack the ability to process confidence metrics of predicted models or to incorporate them into interface-level analyses. ProNA3D addresses this limitation by calculating ipLDDT and iPAE values for individual sub-interfaces, allowing the reliability of predicted contacts to be evaluated directly within the framework used for interface analysis.

Assessing confidence at the sub-interface level revealed that confidence can vary within a complex, with some sub-interfaces modeled more reliably than others, aligning with the pattern observed for protein-protein interfaces (see section 3.1). In ProNA3D, the residue-level pLDDT threshold filter enables users to iteratively narrow interface analysis to the most reliable contacts by applying progressively stricter confidence thresholds. This was demonstrated for the predicted complex between the *Punavirus P1* Cre recombinase and its target DNA (PDB ID: 7RHY, subsection 3.4.4). The reliability of predicted interfaces is practically relevant, since structure-guided applications such as ligand docking or targeted mutagenesis require the identification of interface residues that accurately capture native contact geometry. By enabling the identification of the high-confidence sub-interfaces, ProNA3D ensures that downstream analyses are guided by reliable predictions.

Analysis of the predicted human Fab-antibody-HIV-1 RNA complex (PDB ID: 9C75) further demonstrated that global structural accuracy and local interface quality can vary significantly in predicted models. In this complex, the predicted global orientation of the RNA chain differed from the experimentally determined structure, yet the interface was predicted with high accuracy. This finding highlights that global confidence metrics, which consider all atom pairs, do not necessarily capture the correctness of inter-chain contacts. It therefore highlights the importance of interface-specific quality assessment, particularly for applications that depend on accurate modeling of intermolecular interactions rather than overall chain geometry.

4.4.3 Interface-based density zoning for local quality assessment in cryo-EM structures

Structures of large nucleic acid-containing assemblies are often determined by cryo-EM. A persistent challenge in cryo-EM structure determination is the variation in local

map quality caused by differences in particle mobility, occupancy, and conformational heterogeneity within the sample (Mulvaney et al. 2026). Because global resolution values average these local variations into a single measure, deposited atomic models may contain regions whose accuracy is lower than the reported overall resolution. This limitation is particularly important for interface-forming residues, where atomic details are essential for functional interpretation. By zoning the density volume specifically around each detected sub-interface, ProNA3D enables local map quality to be examined directly in these contact regions.

The cryo-EM structure of the ATP-dependent helicase DHX36 from *Bos taurus* in complex with human G-quadruplex RNA (PDB ID: 8VV2, EMD-43547) illustrated the importance of sub-interface-focused density assessment (see subsection 3.4.5). ProNA3D identified two sub-interfaces within the protein-RNA complex that differed substantially in local map quality. One sub-interface was located in a region of high local resolution, allowing confident interpretation of individual side chains, whereas the other sub-interface was detected in an area of substantially lower local resolution. In this sub-interface, two lysine residues, Lys854 and Lys855, were only partially modeled. Despite their incomplete representation in the atomic model, both residues were identified as RNA-binding contacts by ProNA3D, in agreement with the analysis reported by Banco et al. (2025).

This example also highlights the correspondence between experimental map quality and computational confidence assessment. In the AlphaFold3 prediction of the complex, the residues forming the lower-resolution sub-interface received lower pLDDT scores, reflecting the experimental quality differences without access to the cryo-EM data directly. The agreement between the local resolution from experimental reconstruction and AlphaFold3-based confidence metrics indicated that lower pLDDT values in predicted protein-nucleic complexes may be indicators of increased structural flexibility in the examined area. This aligned with observations from the analysis of protein-protein interfaces (see subsection 4.3.1), which link increased interface B-factors to reduced prediction reliability. For cryo-EM analyses of nucleic acid-containing assemblies, this correspondence suggests that AlphaFold3 confidence metrics may serve as initial indicators of interface regions likely to exhibit reduced local resolution, thereby supporting experimental design and model validation.

4.5 Systematic analysis of nucleic acid-containing interfaces in the PDB

The analysis of 6,047 high-resolution nucleic acid-containing complexes from the PDB, yielding 45,244 sub-interfaces across 5 interaction types, provided a quantitative characterization of structural properties of nucleic acid-containing interfaces. This systematic analysis complemented the large-scale analysis of protein-protein interfaces performed with PICKLUSTER (see section 4.2) and extended it to the full range of macromolecular interaction classes involving nucleic acids.

4.5.1 Interface size distributions across nucleic acid-containing complex classes

Interface size distributions across nucleic acid-containing complexes differed markedly between nucleic acid-only and protein-nucleic acid interaction classes, reflecting the characteristic structural properties of each class and suggesting that the interface size depends substantially on the involved molecule type (see subsection 3.5.1).

Among nucleic acid-only interactions, RNA-RNA interfaces were substantially smaller than DNA-DNA interfaces, aligning with the increased structural flexibility of RNA. RNA can adapt to a wide range of folds, incorporating loops, bulges, and non-canonical base pairs that constrain the size and geometry of inter-chain contacts (Hermann and Patel 1999; Olson et al. 2019). DNA-DNA interfaces, in contrast, often form extended double-helical structures that are generally dominated by Watson-Crick base pairing, although non-canonical conformations can also occur (Neidle 2021; Bansal et al. 2022). RNA-DNA interfaces showed an intermediate size relative to RNA-RNA and DNA-DNA interfaces, suggesting regular geometry imposed by base pairing interactions and stacking, but also indicated the existence of non-standard complex formation.

Protein-nucleic acid interfaces exhibited pronounced asymmetry, with the protein side comprising more contacting residues than the nucleic acid side contributed nucleotides. This pattern likely reflected fundamental differences in molecular geometry. Protein residues are comparatively smaller, and most contain side chains with multiple rotatable bonds, enabling diverse spatial contact geometries with nucleotides, whereas nucleotides are more constrained by the geometry of the nucleic acid backbone and by base stacking (Arakawa et al. 2025). As a result, multiple protein residues can contact a single nucleotide from different directions, creating the observed asymmetry. These results were consistent with the findings reported by Nadassy et al. (1999). The broad range of protein-RNA interface sizes, which exceeds the distribution of protein-DNA, likely reflects the functional diversity of protein-RNA complexes, ranging from small regulatory complexes to very large ribonucleoprotein assemblies such as the ribosome (Draper 1995; Staley and Woolford 2009).

The outlier complexes in the protein-RNA size distribution likely represented structured deviations rather than random variation. The complexes that deviated most from the overall size relationship, including CRISPR-Cas effector proteins bound to crRNA molecules, large ribosomal proteins associated with 25S and 28S rRNA, and Cas9 in complex with guide RNA, corresponded to interaction modes characterized by specific biological functions. This indicated that interface size ratios may capture functional information in addition to geometric characteristics.

4.5.2 Protein binding stabilizes nucleic acid interface flexibility

Across all interface classes, protein binding was associated with reduced flexibility of nucleic acid interface regions (see subsection 3.5.2). RNA-RNA and DNA-DNA interfaces showed consistently higher normalized B-factors than protein-RNA and protein-DNA interfaces, whereas RNA-DNA interfaces exhibited the lowest overall flexibility. This pattern was consistent with previous observations that the DNA backbone can be stabilized through interactions with proteins, which are known to impose structural constraints via backbone contacts, electrostatics, and shape complementarity at the interface (Simon and Gehrke 2009; Schneider et al. 2014).

In contrast, unbound nucleic-acid backbones exhibit substantially higher flexibility. Despite these differences, the standard deviations were similar across interface types, suggesting that heterogeneity in B-factors is a common property of biomolecular interfaces and is not strongly dependent on the interaction type.

The particularly low B-factors observed for RNA-DNA interfaces were notable, as they contrast with previous studies indicating that such hybrids are more flexible than RNA duplexes (Noy et al. 2005). However, given the relatively small number of RNA-DNA sub-interfaces in the dataset, even slight biases could influence the observed values. Therefore, any mechanistic explanation for this observation would require further analysis.

More generally, it should be noted that the observed differences were minor, with median B-factors ranging from 0.23 to 0.32. Accordingly, the results should be interpreted as trends rather than absolute differences.

4.5.3 Connectivity patterns in nucleic-acid containing complexes

The maximum connectivity analysis revealed pronounced differences between nucleic acid-only interfaces and protein-nucleic acid interface classes (see subsection 3.5.3). RNA-RNA, RNA-DNA, and DNA-DNA interfaces exhibited the same median maximum connectivity, with no statistically significant differences being detected between these classes. This similarity was notable as these interface types differ substantially in helix geometry and structural context (Cheatham and Kollman 1997; Noy et al. 2005). These results indicated that the general architecture of nucleotides, rather than the specific nucleic acid identity, determined the maximum possible connectivity per nucleotide. Moreover, the relatively uniform connectivity observed for nucleic acid-only interfaces may be explained by the regularity of duplex DNA and the, even more diverse, yet structured tertiary contacts characteristic of RNA (Butcher and Pyle 2011; Tjong et al. 2014). RNA-RNA interfaces exhibited a broader connectivity distribution compared to the DNA-DNA and DNA-RNA interfaces. This likely reflected the greater structural flexibility of RNA, which can engage in non-canonical interactions and tertiary contacts beyond the constraints of a canonical duplex (Nagaswamy et al. 2002; Lu et al. 2015).

A pronounced local maximum at a connectivity value of 1 was observed in the DNA-DNA connectivity distribution. This pattern likely reflected nicked DNA structures, in which a discontinuity in the backbone is stabilized by base stacking, producing a single inter-chain contact (Aymami et al. 1990). The ability to detect this structural pattern at the level of connectivity highlighted the sensitivity of the interface analysis with ProNA3D.

Protein-nucleic acid interfaces exhibited substantially higher maximum connectivity values and substantially broader distributions than nucleic-acid-only interfaces. This may be explained by the geometric asymmetry between proteins and nucleic acids. Individual nucleotides can simultaneously be in contact with several smaller protein residues, resulting in increased connectivity values. The broader connectivity distributions of protein-nucleic acid interfaces likely reflected the functional diversity of protein-mediated recognition of nucleic acids, which ranges from small regulatory complexes to very large ribonucleoprotein assemblies (Draper 1995; Staley and Woolford 2009; Bellone et al. 2025).

A subset of protein-nucleic acid interfaces exhibited exceptionally high maximum connectivity values, representing specialized binding mechanisms. The majority

corresponded to enzyme-DNA complexes, particularly DNA methyltransferases, in which individual nucleotides are flipped out of the strand and enclosed by protein residues (Neely et al. 2005). A similar pattern was identified in the CRISPR-Cas system component Csm4 in complex with crRNA. The recurrence of nucleotide flipping across complexes from different species, which is associated with the observed extreme connectivity values, suggested that local connectivity may serve as an indicator for functionally specialized interaction modes.

Per-residue connectivity analyses showed that positively charged residues were enriched at high-connectivity positions in protein-nucleic acid interfaces (see subsection 3.5.4). Arginine exhibited the highest connectivity at both protein-RNA and protein-DNA interfaces, followed by lysine and histidine, whereas hydrophobic residues consistently showed low connectivity values, in agreement with previous observations (Krüger et al. 2018). This was consistent with the electrostatic charge complementarity between positively charged side chains and the negatively charged phosphate backbone of the nucleotides and highlighted arginine and lysine as the residues most likely to form contacts at the protein-nucleic acid interface (Cherstvy 2009; Rohs et al. 2010). Occurrence distributions reflected this trend, with arginine and lysine being the most prevalent residue types at both protein-DNA and protein-RNA interfaces, whereas hydrophobic residues were underrepresented. For nucleotides, DNA-DNA interfaces showed a balanced composition that potentially reflected canonical base pairing, whereas RNA-containing interfaces displayed a slight enrichment of purines.

On the nucleotide side, the enrichment of adenine and thymine at protein-DNA interfaces was consistent with preferential protein binding to AT-rich DNA regions, which exhibit increased electronegative potential compared with GC-rich segments (Yesudhas et al. 2017).

4.5.4 Structural features of viral nucleic acid-containing interfaces

Analysis of a viral subset revealed differences compared to the full dataset, which were most pronounced for RNA-containing complexes. In protein-RNA interfaces, the protein-side contribution was larger than in the full dataset, whereas the RNA side exhibited only a slight increase (see section 3.5.6). The asymmetry at protein-RNA interfaces likely reflected the involvement of viral capsid proteins, nucleoproteins, and RNA-dependent RNA polymerases, which form large interfaces with RNA (Arnez and Cavarelli 1997; Klumpp et al. 1997; López-Muñoz and Yewdell 2024). RNA-RNA sub-interfaces were also enlarged and displayed a narrower size distribution, consistent with extended double-stranded RNA in dsRNA viral genomes (Lipfert et al. 2014; Q. Zhang et al. 2023).

In addition, viral nucleic acid-only interfaces showed elevated normalized B-factors relative to the full dataset, with the largest increases observed for RNA-RNA and RNA-DNA interfaces (see section 3.5.6). Interestingly, protein-nucleic acid interfaces in the viral subset showed highly similar normalized B-factor distributions compared to the full dataset, indicating that the stabilizing effect of protein binding on nucleic acid flexibility is conserved regardless of the origin (Schneider et al. 2014). Consequently, the differences in B-factors between nucleic acid-only and protein-nucleic acid interfaces were increased in the viral subset compared to the full dataset. An important consideration in this analysis is the potential influence of crystallographic resolution on B-factors. However, the mean and median resolutions were highly similar between

the full dataset and the viral subset (2.3 Å vs. 2.4 Å). Moreover, only weak correlations were observed between resolution and normalized B-factor, with PCCs of 0.32 for the full dataset and 0.26 for the viral subset. The elevated B-factors in the viral nucleic acid-only interfaces are therefore unlikely to reflect a systematic resolution bias. Instead, this pattern may reflect the conformational demands on viral genomes, which must undergo extensive structural remodeling during replication, translation, and packaging (Simon and Gehrke 2009; López-Muñoz and Yewdell 2024). The observations for RNA-DNA interfaces should be interpreted with caution, due to the small number of sub-interfaces available for analysis for this interface type.

The connectivity analysis of the viral subset revealed interface-type-specific characteristics. Viral nucleic acid-only interfaces displayed connectivity distributions that were highly similar to those of the full dataset, consistent with a nucleotide geometry that imposes geometrical constraints on their connectivity independent of biological context (see section 3.5.6). In contrast, viral protein-nucleic acid interfaces showed increased connectivity compared to the full dataset for both protein-RNA and protein-DNA interfaces. In combination with the enlarged protein-side contribution in protein-RNA interfaces, this increase suggested that proteins in viral complexes formed larger, highly connected interfaces with nucleic acids, aligning with the need for extensive and stable protein-RNA contacts in virus-specific processes, such as RNA-encapsidation and the assembly of ribonucleoprotein complexes (Arnez and Cavarelli 1997; Klumpp et al. 1997; López-Muñoz and Yewdell 2024).

4.6 Limitations

The three tools developed in this thesis (PICKLUSTER, C2Qscore, and ProNA3D) and the systematic evaluation of AlphaFold-derived scoring metrics each involve specific methodological choices that enable their analyses but also limit the range of structures, interaction types, and biological contexts for which their outputs can be reliably interpreted. These limitations do not compromise the results presented, but they should be considered when applying the tools to new systems or extending their use beyond the cases evaluated in this work. The following subsections discuss the main limitations of the tools and workflows described in this thesis.

4.6.1 Limitations of the PICKLUSTER and ProNA3D algorithm

PICKLUSTER and ProNA3D identify interface residues and nucleotides solely based on heavy-atom distance, without considering the BSA, electrostatic character, or the geometry of specific interactions. Heavy-atom pairs within the distance threshold and the corresponding residues and nucleotides are reported as interaction partners, regardless of whether the local chemical environment supports a specific non-covalent interaction. As a result, some reported contacts may not correspond to real interactions. Accordingly, for systematic analyses that require discrimination between specific non-covalent interactions, such as hydrogen bonds or salt bridges, a specialized analysis method should be used in combination with PICKLUSTER and ProNA3D.

Non-standard and post-translationally modified residues are currently excluded from the interface detection algorithm implemented in PICKLUSTER and ProNA3D. As a result, for complexes in which interface-forming residues and nucleotides are

modified, the reported interface will be incomplete. This limitation is particularly relevant for complexes in which post-translational modifications are frequently located at or near the interface, such as signaling complexes and chromatin-associated proteins (Suskiewicz 2024). In addition, neither tool detects solvent-mediated or metal ion-bridged interactions, nor small-molecule binding. Consequently, interfaces that rely substantially on such interactions may be incompletely represented.

The default distance cutoff used for centroid-based clustering of interface residues and nucleotides into sub-interfaces ($d_C = 5.0 \text{ \AA}$) was selected based on the heavy-atom distance threshold and may not be optimal for all complex types or interface geometries. If the cutoff is overly restrictive, a geometrically and physicochemically continuous interface may be divided into multiple sub-interfaces. Conversely, if the cutoff is overly permissive, spatially and physicochemically distinct interface regions may be merged into a single entity. Although the selected default value for d_C appears to yield biologically interpretable results across a broad range of biomolecular complexes, it should ideally be benchmarked against additional interface characteristics, such as electrostatic properties, to further assess whether the resulting sub-interface definitions are biochemically meaningful.

In addition, the confidence metrics provided by PICKLUSTER and ProNA3D, including ipLDDT and iPAE, are currently implemented only for AlphaFold2 and AlphaFold3 predictions. Models generated by other prediction tools that report confidence in different formats are therefore not supported.

The secondary-structure topology plots in ProNA3D are based on base-pair annotations computed by DSSR (Defining the Secondary Structures of RNA). Although DSSR can detect a broad range of nucleic acid geometries (Lu et al. 2015), it does not determine protein secondary structure. Therefore, the topology plots are limited to nucleic acid secondary structure, while proteins are represented only as contacts. Future development could address this limitation by integrating DSSP (Dictionary of Secondary Structure in Proteins) (Kabsch and Sander 1983; Hekkelman et al. 2025), which assigns secondary-structure elements to protein residues based on atomic coordinates, thereby enabling a unified topology representation for both nucleic acid and protein components. In addition, DSSR requires a license, which restricts access to this feature to licensed users.

4.6.2 C2Qscore

The evaluation of the individual scoring metrics (ipLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF) as well as C2Qscore was performed exclusively on heterodimeric protein complexes. Therefore, their performance has not been assessed for homo-oligomeric assemblies, multi-chain complexes, or antibody-antigen interactions, which were explicitly excluded from the benchmark set. Homomeric interfaces are generally predicted with higher accuracy by AlphaFold2 (Evans et al. 2022; Zhu et al. 2023), which may affect the discriminatory power of the confidence metrics. Accordingly, the application of the derived cutoffs to complex types not represented in the benchmark set should be performed with caution and would ideally be supported by recalibration against a suitable reference dataset.

The scoring metrics were evaluated for their ability to assess the structural quality of a predicted interaction between two chains, given the assumption that those chains form a complex. They are not designed to determine whether two chains

interact. Consequently, the classification thresholds derived from ROC analysis are not applicable to protein-protein interaction detection tasks, which require distinguishing true inter-chain contacts from false ones. However, there have been efforts to address this problem by developing specialized scoring approaches (Bryant et al. 2022; Burke et al. 2023; Schweke et al. 2023).

The regression weights of C2Qscore are specific to CF-T, CF-F, and AF3 and should not be applied across different prediction methods. The difference in the relative contribution of ipLDDT to C2Qscore between AF3 and ColabFold-based predictions indicates that the relative importance of each input metric depends on the architecture and training data of the prediction method. Applying the score to results from new, updated, or substantially different prediction methods would therefore require recalibration of the regression weights.

A limitation of the analysis is the discrepancy between the asymmetric unit (AU) and the biological assembly (BA) in a subset of PDB entries. Six benchmark targets were identified in which the annotated dimeric BA differed from the AU, causing alignment failures and artificially low DockQ values that did not reflect the true prediction quality. These cases were excluded from the final benchmark set. However, as the AU may reflect crystallographic packing rather than a biologically relevant state, it would be important to assess whether prediction methods preferentially model the AU or the BA in such cases.

Although the assessment of interface-specific scoring metrics (except for pTM) is the main focus of the analysis described in this thesis, this also introduces a limitation. While such metrics are well-suited to assess local interface accuracy, they do not fully capture the global arrangement of the complex. In particular, an interface may be modeled correctly while the global orientation of the interacting chains remains incorrect, as seen in subsection 3.4.3. This limitation highlights that scoring metrics should not be interpreted in isolation. Instead, a more comprehensive assessment of model quality requires the combined use of multiple scoring metrics, including those that reflect global structural confidence. Measures such as the PAE matrix or pLDDT distributions across the entire complex provide complementary information revealing inconsistencies between local interface accuracy and overall chain orientations.

The use of DockQ as the ground-truth quality measure introduces an additional limitation. DockQ was originally developed for evaluating protein-protein docking predictions, and its CAPRI-derived classification thresholds are based on performance assessments on the MOAL set, which comprises 56,015 models across 118 targets (Moal et al. 2013; Basu and Wallner 2016). Therefore, although these thresholds are well established and widely used in the field, alternative definitions of reference metrics or different cutoff choices could affect the optimal threshold of the evaluated prediction-based scoring metrics.

Moreover, benchmark set targets with structural homologs in the AlphaFold2 and AlphaFold3 training data were intentionally retained to reflect realistic use conditions. However, this may have introduced a bias in favor of these targets, potentially resulting in an overestimation of prediction performance.

4.6.3 Limitations of the large-scale datasets and analyses

Certain characteristics of the datasets used for the large-scale analyses with PICKLUSTER and ProNA3D limit the generalizability of the reported characteristics. The restriction

to high-resolution (≤ 3.0 Å) structures and the dominance of X-ray crystallographic entries in the PDB introduce a bias toward stable, well-ordered complexes. Consequently, structures of complexes that are difficult to resolve by high-resolution structural methods, including transient complexes, complexes involving IDRs, and membrane protein complexes, are likely underrepresented in the analysis.

Although sequence redundancy in the datasets was reduced using a 30% identity threshold for proteins and a 90% threshold for nucleic acids, structural redundancy at the interface level was not considered separately. As a result, conserved contact patterns among structurally similar but sequence-diverse complexes may still be overrepresented in the final distributions.

4.6.4 General considerations

A shared limitation of all three described approaches is that analyses are performed on static three-dimensional structures. Protein-protein and protein-nucleic acid interfaces are inherently dynamic, and their functional properties, such as binding affinity, selectivity, and sensitivity to small-molecule disruption, depend on conformational flexibility that cannot be fully captured by a single static structure (Lill 2011; Tuffery and Derreumaux 2012). Integrating molecular dynamics simulations would therefore provide a more comprehensive characterization of interface dynamics and remains an important direction for future work.

The interaction hotspots identified by PICKLUSTER and ProNA3D are computational predictions that were not experimentally validated in this work. Establishing their functional relevance would require additional confirmation, for example, through alanine-scanning mutagenesis or biophysical assays that assess whether disruption of these residues or nucleotides affects binding. Moreover, neither of the tools estimates the energetic contribution of individual contacts to complex stability. To relate hotspot identification more directly to functional significance, integration with energy-based approaches would be beneficial (Qiu et al. 2018).

5

Conclusions and Outlook

Contents

5.1	Quality assessment of predicted protein-nucleic acid and nucleic acid-only complexes	105
5.2	Nucleic acid-containing and protein-protein complexes as targets for structure-guided drug design	106
5.3	Interface property profiles for molecular class annotation	106
5.4	Structural determinants of prediction quality in nucleic acid-containing complexes	107

This thesis focused on three interconnected aims, resulting in the development of validated computational methods and new insights into the organization of biomolecular interfaces.

The first aim of this thesis was to address the lack of spatial decomposition approaches in existing protein interface analysis tools and to investigate how sub-interface-level analysis can support both detailed characterization of individual complexes and large-scale studies of interface organization across the PDB. In addition, this work aimed to improve the interpretability of AlphaFold-based predictions by integrating interface-specific confidence metrics into the analysis. These objectives were achieved through the development of PICKLUSTER (Genz et al. 2023), which provides a framework for the identification, subdivision, and confidence-aware analysis of protein-protein interfaces in both experimentally determined and computationally predicted structures. Application of PICKLUSTER to a predicted *Epstein-Barr virus* protein complex demonstrated that individual sub-interfaces can differ substantially in prediction confidence, highlighting the importance of sub-interface-level analysis for prioritizing regions in downstream applications such as structure-guided drug discovery. Furthermore, large-scale analysis of 19,821 protein complex structures from the PDB revealed characteristic patterns in the size distribution and spatial organization of protein-protein sub-interfaces across the structural proteome, including extreme cases such as a complex from *Porphyridium purpureum* containing 4,844 sub-interfaces.

The second aim of this thesis was to systematically evaluate confidence metrics for AI-predicted protein complex models, derive prediction method- and metric-specific classification thresholds, and determine whether combined scoring approaches provide more robust quality assessment than individual metrics alone. These objectives were investigated through a benchmarking study of 223 heterodimeric protein complexes predicted using ColabFold with templates, ColabFold without templates, and AlphaFold3, in which commonly used confidence metrics, including pLDDT, iPAE, pTM, ipTM, *model confidence*, pDockQ2, and VoroIF, were systematically evaluated (Genz et al. 2025a). Across all three prediction methods, ipTM and *model confidence* were consistently identified as the most reliable metrics for interface quality assessment. Prediction method- and metric-specific binary classification thresholds were derived, providing a practical framework for distinguishing correct from incorrect predictions. Finally, C2Qscore was developed as a combined scoring approach to improve robustness beyond individual metrics and was subsequently validated on independent X-ray crystallography- and cryo-EM-derived test sets.

The third aim of this thesis was to address limitations of existing analysis tools for nucleic acid-containing complexes, which are often fragmented and lack integrated support for confidence-aware sub-interface analysis, secondary-structure topology visualization, and interactive graphical analysis across diverse macromolecular interaction types. In addition, this work aimed to investigate whether spatial sub-interface decomposition is equally informative for protein-nucleic acid and nucleic acid-only complexes as for protein-protein interfaces. These objectives were addressed by developing ProNA3D (Genz and Topf 2026), which extends the PICKLUSTER framework to nucleic acid-containing complexes and integrates sub-interface identification, two-dimensional topology analysis, cryo-EM density zoning, and confidence metric visualization. Application of ProNA3D to a diverse range of protein-nucleic acid and nucleic acid-only complexes established sub-interface decomposition as equally informative for these interaction classes. Connectivity-based topology analysis identified interaction hubs with potential relevance to complex stability, including nucleotide A38 as a central hub in a dimeric hammerhead ribozyme complex and nucleotide A52 in a human Fab antibody-HIV-1 RNA complex. In addition, the cryo-EM density zoning feature supported localized map validation for an RNA helicase complex from *Bos taurus*, while integration of AlphaFold3 confidence metrics at the sub-interface level improved the interpretability of predicted nucleic acid-containing complexes.

The systematic analyses of nucleic acid-containing interfaces across the PDB using ProNA3D revealed structural characteristics relevant to the understanding of macromolecular interface organization. Nucleic acid-only interfaces were characterized by conserved connectivity patterns, whereas protein-nucleic acid interfaces exhibited substantially higher and more heterogeneous connectivity, reflecting the diversity of protein-binding modes. Protein binding consistently reduced the flexibility of nucleic acid interfaces, a property observed across both viral and non-viral complexes. In addition, viral protein-nucleic acid interfaces were characterized by larger size and increased connectivity compared to the overall dataset, potentially providing a structural basis for their recognition as a distinct interface class. Connectivity analysis further identified base-flipping as a mechanism resulting in extreme connectivity in methyltransferase-DNA and Csm4-RNA complexes, highlighting specialized interaction modes with potential relevance for therapeutic targeting.

Building on the framework and results introduced and described in this thesis, several avenues for future work in structural bioinformatics and structure-guided drug discovery emerge.

5.1 Quality assessment of predicted protein-nucleic acid and nucleic acid-only complexes

A methodological gap identified by this work is the absence of a validated scoring approach for the assessment of predicted protein-nucleic acid and nucleic acid-only complexes. C2Qscore and the individual scoring metrics provide an evaluated quality assessment for protein-protein heterodimers predicted with AlphaFold2 and AlphaFold3 (see subsection 4.3.3). However, while confidence metrics such as pLDDT and PAE are also available for protein-nucleic acid and nucleic acid-only complexes, they have not yet been systematically evaluated at scale, mainly due to the significantly lower number of available reference structures in the PDB. Assessment is further complicated by the lack of a widely accepted reference-based scoring metric with classification thresholds, leading different studies to rely on distinct quality scores for model evaluation. In CASP16, for instance, multimeric nucleic acid-containing complexes were ranked using a Z-score-based quality metric (Kretsch et al. 2026). Recent benchmark studies, such as (Cui et al. 2025), have developed datasets for predicted protein-nucleic acid complexes and assessed them using metrics such as RMSD, TM-score, and lDDT. However, these studies have largely focused on the evaluation of prediction methods rather than the scoring approaches, highlighting the need for the development and validation of dedicated scoring methods. The emergence of AlphaFold3 (Abramson et al. 2024) and RoseTTAFoldNA (Baek et al. 2023b) has enabled the AI-based prediction of nucleic acid-containing complexes, increasing the number of available models and highlighting the need for reliable assessment metrics.

Therefore, the development and evaluation of assessment scores for predicted protein-nucleic acid and nucleic acid-only complexes represents an important direction for future work. A promising strategy would be a machine-learning-based scoring function trained on physicochemical and geometric features specific to these interface types. Similar approaches exist for protein-protein complexes like the neural network-based method implemented in VoroIF (Olechnovič and Venclovas 2023), which relies on interface contacts derived from Voronoi tessellation. Such methods require appropriate training data, and the dataset curated for the large-scale analyses with ProNA3D of PDB structures could serve as a basis. As the number of AI-predicted protein-nucleic acid and nucleic acid-only complex structures with corresponding experimental counterparts in the PDB continues to increase, sufficiently large training sets for data-driven scoring approaches are expected to become possible. An important requirement for this is the development of a suitable ground-truth quality metric. DockQ (Basu and Wallner 2016), developed for protein-protein interfaces, cannot be directly applied to nucleic acid-containing interactions. Therefore, a nucleic acid-specific, analogous metric will need to be defined and benchmarked to assess nucleic acid binding accuracy.

5.2 Nucleic acid-containing and protein-protein complexes as targets for structure-guided drug design

Structure-guided drug discovery has historically focused on the active sites of enzymes or ligand-binding sites of receptors and ion channels, but the increasing relevance of protein-protein, protein-nucleic acid, and nucleic acid-only interfaces as therapeutic targets underscores the need for analytical frameworks applicable across complex types (Winter et al. 2012; Haga and Phinney 2023; Genz et al. 2025b). The pipeline presented in this thesis addresses this by enabling interface detection and analysis of protein-protein, protein-nucleic acid, and nucleic acid-only assemblies through PICKLUSTER and ProNA3D. In addition, it provides a guide for the confidence-based quality assessment for protein-protein complex models. The development of a corresponding scoring method for protein-nucleic acid complexes (see section 5.1) represents a key next step toward establishing a unified quality assessment framework for all complex types accessible to AlphaFold3.

The identification of inter-chain interfaces represents a central step in structure-guided drug discovery. Protein-protein and protein-nucleic acid interfaces often consist of multiple spatially distinct sub-interfaces that can differ in size, residue composition, physicochemical character, and structural confidence (Keskin et al. 2005; Lunde et al. 2007). PICKLUSTER and ProNA3D enable the individual evaluation of these sub-interfaces with respect to their potential druggability. Small, high-confidence sub-interfaces provide well-defined contact geometries that may be accessible to small molecules or constrained peptides (Koes and Camacho 2012). For AI-predicted complexes, confidence-based filtering provides an important pre-selection step. Models passing method-specific thresholds can be used as reliable templates, which is particularly important for downstream analyses.

Within high-confidence sub-interfaces of nucleic acid-containing complexes, connectivity analysis using interface graphs or topology plots provided by ProNA3D enables the identification of interaction hotspots. In protein-RNA and protein-DNA interfaces, highly connected arginine and lysine residues often represent positions whose disruption would affect multiple contacts simultaneously, highlighting them as potential targets for inhibition (see subsection 4.5.3). Furthermore, the base-flipping observed in methyltransferase-DNA and Csm4-RNA complexes may represent an interesting mechanism for drug targeting or chemical modifications, as the flipped nucleotide is enclosed by protein residues.

5.3 Interface property profiles for molecular class annotation

The large-scale ProNA3D analysis indicates that nucleic acid-containing complexes with specific functions can potentially be identified based on interface properties such as interface size and maximum connectivity. Ribosomal proteins, CRISPR-Cas effectors, and DNA methyltransferases emerged as distinct outliers in the interface size and connectivity distributions without requiring prior class annotation (see section 4.5, subsection 4.5.3). This suggests that interface property profiles may

enable the functional annotation of uncharacterized complexes and highlight atypical binding modes for further study.

A machine learning-based classifier trained on ProNA3D descriptors, including sub-interface size, maximum connectivity, secondary structure, and confidence metrics, may enable automated class annotation for predicted nucleic acid-containing complexes. This capability is increasingly important as the recent advances in structure prediction have created a growing gap between large-scale structure generation and the slower process of accurate functional annotation (Varadi et al. 2025). Moreover, such an approach may be particularly valuable for viral protein-RNA complexes. ProNA3D-based analyses indicate that these complexes are characterized by enlarged interfaces and elevated connectivity (see subsection 4.5.4), consistent with the presence of capsid proteins, nucleoproteins, and RNA-dependent RNA polymerases (Arnez and Cavarelli 1997; Klumpp et al. 1997; López-Muñoz and Yewdell 2024). A classifier capable of recognizing these features as signatures could therefore identify the specific function of viral complexes without requiring prior biological annotation.

5.4 Structural determinants of prediction quality in nucleic acid-containing complexes

Although AlphaFold3 has significantly improved nucleic acid-containing complex structure prediction, predicting which targets can be modeled reliably remains a key challenge for large-scale structure-based studies. Analysis of the protein-protein benchmark dataset (see subsection 3.3.1) indicates that mispredicted protein-protein heterodimers tend to have smaller and more flexible interfaces, suggesting that these properties may serve as indicators of model accuracy prior to prediction. ProNA3D provides comparable interface descriptors, including sub-interface size, normalized B-factors, and residue connectivity, for nucleic acid-containing complexes. This offers a basis to investigate whether similar relationships between interface properties and prediction accuracy exist for these complex types. Such analyses would enable the development of pre-screening strategies to prioritize targets with favorable structural properties, guiding the selection of suitable systems for structure prediction.

However, this question remains unresolved mainly due to the lack of a suitable benchmark set for nucleic acid-containing complex predictions. As the number of predicted complexes increases, a systematic analysis could establish how interface properties influence prediction accuracy.

6

Methods

Contents

6.1	Tool development: PICKLUSTER	109
6.1.1	Design and implementation	109
6.1.2	Interface calculation and clustering algorithm	110
6.1.3	Additional features and scoring metrics	111
6.1.4	Dataset for the large-scale analysis of protein-protein interfaces	112
6.2	Assessing scoring metrics for protein complex predictions	112
6.2.1	Benchmark set and X-ray crystallography test set construction (Figure 6.1)	113
6.2.2	Model generation	114
6.2.3	Reference-based score: DockQ	115
6.2.4	Prediction-based scores	115
6.2.5	Combined scoring: C2Qscore	115
6.2.6	Pearson correlation coefficient	116
6.2.7	ROC curves and AUC	116
6.2.8	Accuracy	116
6.2.9	Precision	116
6.2.10	Wilcoxon signed-rank test	116
6.3	ProNA3D	117
6.3.1	General functions and implementation	117
6.3.2	Interface Extraction	117
6.3.3	Topology analysis	118
6.3.4	Two-dimensional interaction networks	118
6.3.5	AlphaFold3 scoring	119
6.3.6	Cryo-EM map analysis	119
6.3.7	Dataset curation for large-scale PDB analysis	119
6.3.8	Viral subset	120
6.3.9	B-factor normalization	120
6.3.10	Connectivity analysis	120
6.3.11	Mann-Whitney U test	121
6.4	Data and software availability	121

This chapter describes the computational methods underlying the three contributions of this thesis. The first part covers the design and implementation of PICKLUSTER, including the algorithms for interface detection, spatial clustering, and confidence metric calculation, as well as the curation of the large-scale protein-protein interface dataset (see section 6.1). The second part details the construction of the heterodimeric protein-protein benchmark dataset, the scoring metrics evaluated, and the development and validation of C2Qscore (see section 6.2). The third part describes the implementation of ProNA3D and the methods used for the large-scale analysis of nucleic acid-containing interfaces, including the curation of the viral subset (see section 6.3).

6.1 Tool development: PICKLUSTER

This section describes the design and implementation of PICKLUSTER, a plug-in for the visualization software UCSF ChimeraX (Pettersen et al. 2021) and a command-line tool for large-scale analyses. PICKLUSTER identifies and characterizes protein-protein interfaces. The content of this section is based on Genz et al. (2023).

6.1.1 Design and implementation

PICKLUSTER is implemented as a UCSF ChimeraX (Pettersen et al. 2021) plug-in and was developed for ChimeraX version 1.4. However, the software has been validated for compatibility with ChimeraX releases up to version 1.10. HTML, JavaScript, Qt, and Jinja2 are used for the front-end in PICKLUSTER.

In addition to the graphical plug-in, PICKLUSTER is also available as a command-line tool. This implementation is designed to conduct large-scale or automated analyses without the need for a visualization environment. Because ChimeraX is required for three-dimensional visualization of molecular structures, the command-line version regenerates tabulated results as an Excel spreadsheet (.xlsx).

PICKLUSTER is implemented in Python 3.9. Both the plug-in and the command-line version require as input the structure of a protein complex, either experimentally determined or computationally predicted. Supported input formats include PDB and mmCIF files. The plug-in implementation relies on the internal ChimeraX atomic module for handling molecular structures, whereas the command-line version uses the Gemmi (Wojdyr 2022) library (version 0.5.8) for structure parsing and analysis.

Besides the protein complex structure input file, the interface calculation requires several user-defined parameters:

- **Chain selection:** The user specifies the chain identifiers for which the interface is to be calculated. Alternatively, the calculation can be performed for the entire complex.
- **Interface clustering:** The user can specify whether decomposition into sub-interfaces using the clustering algorithm (see subsection 6.1.2) is performed.
- **Clustering cutoff:** If clustering is enabled, a distance cutoff for the clustering algorithm must be provided (details are described in subsection 6.1.2).

- **Input type:** Three types of input are supported: (i) experimentally determined protein complex structures (Protein Complex Structure), (ii) AlphaFold2 predictions (Protein Complex Model (AlphaFold2)), and (iii) AlphaFold3 predictions (Protein Complex Model (AlphaFold3)).
- **pLDDT-based filtering:** AlphaFold2- and AlphaFold3-predictions require the selection of a pLDDT threshold (default: 50). Residues with pLDDT values below this threshold are excluded from the interface calculation. The per-residue pLDDT is used for this procedure. For AlphaFold3 predictions, this value is determined by averaging corresponding per-atom pLDDT values.
- **AlphaFold score files:** Since AlphaFold2 and AlphaFold3 use different file formats for their confidence scores, the prediction type must be specified. For the calculation and reporting of the quality metrics, PICKLUSTER additionally requires the corresponding JSON or pickle output files generated during the AlphaFold2/3 prediction process (except for ipLDDT).

6.1.2 Interface calculation and clustering algorithm

PICKLUSTER identifies protein-protein interfaces by computing distances between heavy atoms (non-hydrogen atoms) of user-defined chains or across the entire complex. The computation is accelerated using a KDTree data structure implemented in the SciPy (Virtanen et al. 2020) library, enabling efficient identification of contacting atom pairs. Heavy atoms located within a distance threshold (d_{HA}) of 5.0 Å are considered to form an inter-chain contact. The interface I is defined as follows:

$$I = \{(r_i, r_j) \mid \min \{a \in r_i, b \in r_j\} \|x_{a,i} - x_{b,j}\| \leq d_{\text{HA}}\} \quad (6.1)$$

where r_i and r_j are residues belonging to chains i and j , respectively, and a and b represent their corresponding heavy atoms with coordinates x .

For AlphaFold2 and AlphaFold3 predictions, residues with pLDDT values below a user-defined threshold are excluded from the interface determination (see subsection 6.1.1).

Moreover, PICKLUSTER enables the decomposition of interfaces into spatially distinct sub-interfaces. This is implemented via a clustering procedure based on residue proximity. First, the center of mass is computed for each interface residue. When clustering is enabled, pairwise distances between these centers are evaluated. Residues are assigned to the same sub-interface S_k if their centers lie within a user-defined cutoff distance d_C (default: 5.0 Å). The set of sub-interfaces S is defined as:

$$S = \{S_k \subseteq I \mid (r_i, r_j) \in S_k \text{ if } \|c_i - c_j\| \leq d_C\} \quad (6.2)$$

where r_i, r_j are residues in the interface I . The centroids c_i and c_j of residues r_i and r_j on chains i and j are defined as the mean positions of their atoms.

This approach enables the identification of multiple spatially distinct sub-interfaces within a single protein-protein interface.

Alternatively, PICKLUSTER allows visualization of the interface without clustering, allowing the protein-protein interface to be analyzed either as a single entity or as subdivided regions.

6.1.3 Additional features and scoring metrics

Beyond the required input parameters, PICKLUSTER provides an extended set of features that support the analysis and interpretation of protein-protein interfaces. These include visualization tools, detailed interface information, and confidence assessment.

Visualization and coloring options PICKLUSTER provides multiple visualization options to support both structural and sequence-level analysis of identified interfaces. These include the generation of a SCF (Sequence-Coloring Format) file for the detected sub-interfaces. This enables residues within sub-interfaces to be colored in the ChimeraX sequence viewer according to the coloring scheme used in the three-dimensional structure. As a result, sub-interfaces can be examined not only spatially but also in the context of the corresponding sequence regions.

For protein complex structures predicted by AlphaFold2 or AlphaFold3, PICKLUSTER provides the option to color interface residues according to their pLDDT values. This functionality enables visual assessment of the prediction confidence within the interface region, supporting the evaluation of structural reliability.

Output and interaction details PICKLUSTER provides multiple output options that offer different levels of detail for the interface analysis. By default, results are presented in the ChimeraX Log panel as a table listing all identified sub-interfaces and their corresponding residues. These sub-interfaces are interactively linked to the three-dimensional structure, allowing users to select and highlight them via the respective sub-interface identifiers in the table.

For a more comprehensive analysis, atomic-level contact visualization can be enabled, displaying all atomic contacts in the interface as pseudobonds in the three-dimensional structure and creating a dedicated ChimeraX object for further analysis.

Moreover, PICKLUSTER provides residue-level contact information. In this mode, a table is generated listing all residue pairs involved in inter-chain interactions within each sub-interface.

Scoring-metrics for AlphaFold-based predictions PICKLUSTER additionally implements several scoring metrics for the evaluation of AlphaFold2 and AlphaFold3 protein complex predictions on the sub-interface level. These include the interface pLDDT (ipLDDT), which represents an interface-specific version of the pLDDT. It is computed by averaging per-residue pLDDT values (AlphaFold2) or per-atom pLDDT values (AlphaFold3) over all interface residues or atoms, providing a direct measure of confidence for the sub-interface. The ipLDDT is defined as follows:

$$\text{ipLDDT} = \frac{1}{|I|} \sum_{j=1}^{|I|} \begin{cases} \text{pLDDT}_j & (\text{AlphaFold2}) \\ \frac{1}{A_j} \sum_{i=1}^{A_j} \text{pLDDT}_{ij} & (\text{AlphaFold3}) \end{cases} \quad (6.3)$$

where $|I|$ is the number of interface residues, A_j is the number of atoms in residue j , pLDDT_j is the per-residue pLDDT of residue j (AlphaFold2), and pLDDT_{ij} is the pLDDT of atom i in residue j (AlphaFold3).

Additionally, PICKLUSTER supports the visualization of the predicted aligned error (PAE) matrix within the ChimeraX Log, reports the maximum PAE value, and

computes the interface PAE (iPAE). For iPAE calculation, all PAE values corresponding to interacting residue pairs at the interface are extracted from the PAE matrix and averaged, resulting in a single score for each interface or sub-interface. The iPAE is defined as follows:

$$\text{iPAE} = \frac{1}{M} \sum_{i,j=1}^M \left(\frac{\text{PAE}_{i,j} + \text{PAE}_{j,i}}{2} \right) \quad (6.4)$$

where $\text{PAE}_{i,j}$ and $\text{PAE}_{j,i}$ correspond to the entries in the PAE matrix associated with residue pairs in the interface I , and M represents the total number of residue pairs between chains i and j , each contributing two PAE values.

Moreover, PICKLUSTER provides access to further AlphaFold-derived metrics, including the predicted template modeling score (pTM), the interface pTM (ipTM), and the *model confidence* ($0.8 \cdot \text{ipTM} + 0.2 \cdot \text{pTM}$) for the protein complex. Except for ipLDDT, the calculation of these metrics requires the files containing the scores from AlphaFold2/3, as described in subsection 6.1.1.

PICKLUSTER v2.0 additionally integrates VoroIF-GNN (Olechnovič and Venclovas 2023) and C2Qscore (see subsection 6.2.5) to enable combined quality assessment of predicted protein complex models. VoroIF-GNN requires external installation and execution. PICKLUSTER accepts its output file as an additional input for downstream analysis. The computation of C2Qscore relies on both the VoroIF-GNN output and AlphaFold2/3-derived scoring files. Further details on the C2Qscore are provided in subsection 6.2.5.

6.1.4 Dataset for the large-scale analysis of protein-protein interfaces

First, PDB entries were filtered to retain structures containing at least two interacting protein chains with a resolution of $\leq 3.0 \text{ \AA}$ and a release date prior to 23.02.2026. To reduce sequence redundancy, clustering was performed using MMseqs2 (Steinegger and Söding 2017) in a multi-step workflow. Protein sequences corresponding to individual chains were extracted from the SEQRES records of the mmCIF files and clustered at a sequence identity threshold of 30%.

All interfaces within each complex were subsequently identified using PICKLUSTER with default d_{HA} , without application of the clustering algorithm (see subsection 6.1.2). The resulting interfaces were mapped to interactions between sequence clusters, and for each unique cluster pair, a single representative structure was randomly selected.

In addition, a minimum sequence length of 30 residues per chain was required. Application of PICKLUSTER with default d_{HA} and d_{C} parameters (see subsection 6.1.2) to the filtered dataset yielded 49,321 sub-interfaces, corresponding to 29,288 interfaces across 19,821 PDB entries.

6.2 Assessing scoring metrics for protein complex predictions

The following section describes the methods and data curation procedures used for the systematic evaluation of scoring metrics for protein complexes predicted by ColabFold and AlphaFold3. The methodology presented here is based on Genz et al. (2025a).

6.2.1 Benchmark set and X-ray crystallography test set construction (Figure 6.1)

Experimentally determined structures from the Protein Data Bank (PDB) were used for the benchmark dataset. PDB entries were initially filtered to retain only heteromeric complexes comprising 2 protein chains. To reduce potential bias from structures included in the AlphaFold2 or AlphaFold3 training sets, only entries released after September 30, 2021 (the final date of the AlphaFold training data) and before July 11, 2025 were considered. To ensure high structural reliability, only structures with a resolution of ≤ 3.0 Å were included. The dataset was subsequently clustered at a sequence identity threshold of 30% using MMseqs2 (Steinegger and Söding 2017) with default PDB settings to further reduce redundancy. For each resulting cluster, the asymmetric unit (AU) was retrieved from the PDB, yielding an initial set of 684 complexes.

To ensure that the selected entries contained interacting protein chains, PICKLUSTER was applied to identify interfaces. Complexes were classified as interacting if at least three residues from each chain formed contacts with the partner chain within a d_{HA} of 5.0 Å. Non-interacting assemblies were excluded from the dataset. Only complexes in which both chains comprised more than 50 residues were retained. These filtering steps reduced the dataset to 389 protein complexes.

Subsequently, antibody-antigen complexes were excluded, as they constitute a distinct class of interactions characterized by hypervariable loops shaped by immune-specific diversification mechanisms (Davies et al. 1990; Janeway et al. 2001). These features introduce functional constraints beyond sequence similarity. Previous studies have shown that AlphaFold-based methods exhibit reduced accuracy in modeling such complexes (Evans et al. 2022; Yin et al. 2022). Entries containing antibodies or nanobodies were identified and removed using the SAbDab database (Dunbar et al. 2014). To ensure that the remaining complexes were heterodimeric, intra-complex sequence homology was assessed using CD-HIT with a sequence identity cutoff of 30% (Fu et al. 2012). After these filtering steps, the dataset comprised 268 complexes.

Six complexes were identified in which the asymmetric unit (AU) differed from the biological assembly (BA) annotation in the PDB. In these cases, predicted models resembled only one of the two representations, leading to alignment errors and artificially low DockQ scores. These complexes were therefore excluded from the benchmark. A subset of 39 heterodimeric complexes was randomly withheld from the benchmark to serve as an independent test set (*X-ray test set*). The final benchmark dataset comprised 223 heterodimeric protein complexes.

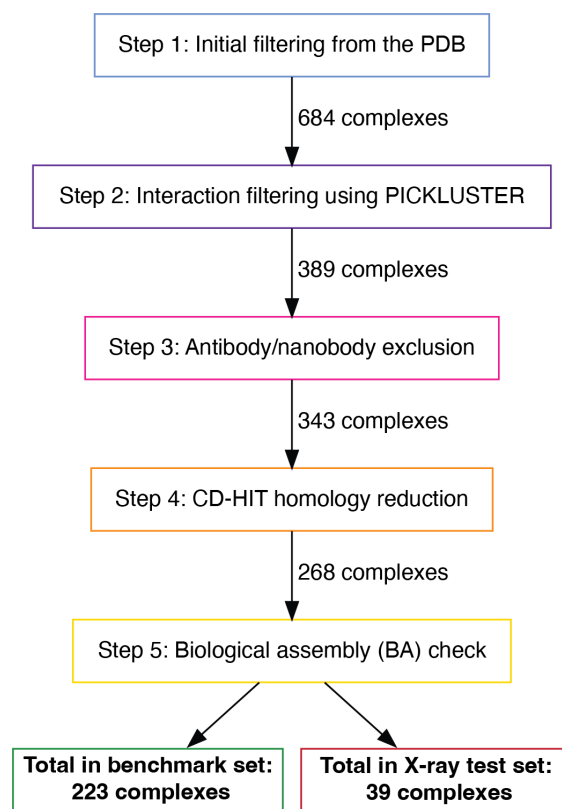


Figure 6.1: Curation pipeline for the benchmark and the X-ray test set. Adapted from Genz et al. (2025a).

Large assembly test set. To evaluate the generalizability of the scoring metrics and derived classification cutoffs in a more complex biological context, an additional test dataset was constructed from dimers extracted from higher-order assemblies resolved by cryo-EM (large assembly test set). The curation procedure followed that of the benchmark set (Figure 6.1), with modifications. Structures were required to originate from trimeric or higher-stoichiometry complexes with a resolution of ≤ 2.3 Å and a release date after September 1, 2023. The stricter resolution cutoff accounts for the variability in local resolution characteristic to cryo-EM data. Intra-dataset redundancy was reduced using CD-HIT at a 30% sequence identity threshold, yielding a final dataset of 156 dimeric target structures.

6.2.2 Model generation

Three AI-based structure prediction approaches were evaluated, resulting in three corresponding prediction sets. For the CF-F predictions, structural models were generated from sequence inputs using local ColabFold v1.5.2, which is based on AlphaFold v2.3 (Mirdita et al. 2022). MSAs were constructed using the `mmseqs_uniref_env` mode, and three recycling steps were applied. For each target, five models were generated and relaxed (Hornak et al. 2006).

For the CF-T predictions, models were generated using the same protocol as for CF-F, with access to template structures from the PDB. To avoid data leakage, the template database was restricted to entries released on or before September 30, 2021, thereby excluding all benchmark set targets from template selection.

The AF3 predictions were generated using the AlphaFold3 web server (Abramson et al. 2024) with default settings, including the use of templates released prior to September 30, 2021.

For all three approaches, input sequences were derived directly from the corresponding PDB sequence records.

6.2.3 Reference-based score: DockQ

The quality of predicted protein complex models was assessed by comparison to the corresponding experimental reference structure using DockQ v2.0 (Basu and Wallner 2016; Mirabello and Wallner 2024). DockQ combines multiple CAPRI evaluation criteria into a single continuous score ranging from 0 to 1, with increased values indicating higher agreement between predicted and native interfaces (Lensink and Wodak 2013). Predictions were classified into four quality categories based on CAPRI thresholds: incorrect (DockQ < 0.23), acceptable (≥ 0.23), medium quality (≥ 0.49), and high quality (≥ 0.80).

6.2.4 Prediction-based scores

Prediction-based scores were evaluated for their ability to assess predicted model quality in the absence of an experimental reference structure, including pLDDT (Jumper et al. 2021), PAE (Evans et al. 2022), ipLDDT (Jumper et al. 2021; Genz et al. 2023), iPAE (Evans et al. 2022; Genz et al. 2023), pTM (Jumper et al. 2021), ipTM (Evans et al. 2022), *model confidence* (Evans et al. 2022), pDockQ2 (Zhu et al. 2023), and VoroIF-GNN (Olechnovič and Venclovas 2023). A detailed description of each score, including its numerical range, the structural property it captures, and whether it is a global or interface-specific score, is provided in Table A.1.

6.2.5 Combined scoring: C2Qscore

To develop a more robust scoring approach, unweighted combinations as well as linear and Lasso regression models of the individual prediction-based scoring metrics were evaluated. Before regression was applied, ipLDDT and iPAE were rescaled to the interval [0, 1], where a value of 1 corresponds to optimal model quality. As *model confidence* and pDockQ2 represented combined scores themselves, they were excluded to avoid redundancy. The remaining five metrics (ipLDDT, iPAE, pTM, ipTM, and VoroIF) were used as input features for the combination approaches.

In an initial step, all possible combinations of scores (without repetition) were evaluated using equal weighting, and the subset achieving the highest PCC with DockQ was identified. Subsequently, linear regression models were fitted both to the full set of five scores and to the best-performing subset. In addition, Lasso regression was assessed as an alternative approach, promoting model simplification by penalizing non-informative features. All regression models were implemented using the `linear_model` module of `scikit-learn` (version 1.3.0) (Pedregosa et al. 2011).

6.2.6 Pearson correlation coefficient

The Pearson correlation coefficient (PCC) was used to quantify the linear relationship between each prediction-based score and the reference DockQ value. The PCC ranges from -1 to $+1$, where a value of 0 indicates no linear correlation and values of ± 1 correspond to a perfect linear relationship. Positive values indicate a direct correlation, whereas negative values reflect an inverse relationship. PCC values were computed using the stats module of SciPy v1.11.3 (Virtanen et al. 2020).

6.2.7 ROC curves and AUC

Receiver operating characteristic (ROC) curve analysis was performed to determine optimal classification thresholds for each prediction-based score. For each score, the true positive rate (TPR) and false positive rate (FPR) were calculated incrementally increasing threshold values and used to construct ROC curves. Thresholds were adjusted in steps of 5 for pLDDT and ipLDDT, in steps of 1 for PAE and iPAE, and in steps of 0.05 for all remaining scores. The optimal cutoff was defined as the point with the maximum Euclidean distance from the diagonal (Melo and Sali 2007). The area under the ROC curve (AUC) was computed as a measure of discriminatory performance, where a value of 0.5 indicates random classification, and 1.0 indicates optimal separation. ROC curves and AUC values were calculated using the metrics module of scikit-learn v1.3.0 (Pedregosa et al. 2011).

6.2.8 Accuracy

Classification accuracy was defined as the ratio $(TP + TN) / (TP + TN + FP + FN)$, where TP , TN , FP , and FN correspond to the numbers of true positives, true negatives, false positives, and false negatives, respectively (Powers 2020). It quantifies the proportion of correctly classified instances. The metric ranges from 0 to 1, with higher values indicating a greater proportion of correct predictions.

6.2.9 Precision

Precision was defined as $TP / (TP + FP)$ and represents the proportion of correctly predicted positive instances among all predictions classified as positive. It quantifies the ability of a model to minimize the number of false positive predictions. The metric ranges from 0 to 1, where 1 indicates that all predicted positives are correct, and lower values indicate an increasing number of false positives.

6.2.10 Wilcoxon signed-rank test

The Wilcoxon signed-rank test is a nonparametric procedure for evaluating differences in paired observations that does not assume a normal distribution of the underlying data. It was used to assess whether observed differences in classification performance between prediction-based scoring metrics were statistically significant. A significance threshold of $\alpha = 0.05$ was applied. The test was implemented using the stats module of SciPy (version 1.11.3) (Virtanen et al. 2020).

6.3 ProNA3D

This section presents the design and implementation of ProNA3D, which is available both as a plug-in for UCSF ChimeraX (Pettersen et al. 2021) and as a command-line tool for large-scale analyses. ProNA3D builds on the interface detection algorithm established in PICKLUSTER but extends it substantially in both scope and functionality. In contrast to PICKLUSTER, which is limited to protein-protein interfaces, ProNA3D enables the identification of interfaces across a wide range of macromolecular assemblies, including protein-RNA, protein-DNA, RNA-RNA, RNA-DNA, and DNA-DNA complexes. Parts of the methodology presented here are based on Genz and Topf (2026).

6.3.1 General functions and implementation

ProNA3D is implemented as a plug-in for UCSF ChimeraX 1.10 (Pettersen et al. 2021). The graphical user interface is built using HTML, JavaScript, Qt, and Jinja2, providing an interactive and user-friendly environment within ChimeraX. In addition to the graphical plug-in, ProNA3D is available as a command-line tool for large-scale analyses that do not require visualization. The command-line version generates tabulated results as Excel spreadsheets (.xlsx). Both ProNA3D versions are implemented in Python 3.9

As PICKLUSTER, ProNA3D requires a complex structure as input (see subsection 6.1.1), which may be derived from experimental data or computationally predicted. Supported input formats include PDB and mmCIF files. The plug-in implementation relies on the internal ChimeraX `atomic` module for handling molecular structures, whereas the command-line version uses the `Gemmi` (version 0.7.4) (Wojdyr 2022) library for structure parsing and analysis.

The input parameters required for ProNA3D are largely analogous to those described for PICKLUSTER (see subsection 6.1.1), with additional options.

In addition to identifying interfaces between user-defined chain pairs or across the entire complex, ProNA3D enables the calculation of interfaces for a single chain against all other chains in the complex. This option is particularly useful when the analysis focuses on a single chain and its contacts with all other components of the complex.

Furthermore, the plug-in supports the export of results for downstream analysis. A CSV file summarizing all detected sub-interfaces can be generated, facilitating further processing or visualization using external software tools.

6.3.2 Interface Extraction

The interface detection algorithm implemented in ProNA3D is adapted from PICKLUSTER (see subsection 6.1.2). As in PICKLUSTER, interacting residues or nucleotides are identified based on heavy-atom distances (d_{HA}), with computational efficiency achieved by using a KDTree data structure (Equation 6.1). In contrast to PICKLUSTER, ProNA3D allows the definition of the d_{HA} cutoff for interface detection. If no d_{HA} is specified, a default value of 5.0 Å is applied.

As in PICKLUSTER (see subsection 6.1.2), ProNA3D allows for decomposition of interfaces into spatially distinct sub-interfaces. When enabled, interacting residues or nucleotides are grouped into sub-interfaces using an approach conceptually equivalent to that implemented in PICKLUSTER (Equation 6.2). The default clustering cutoff (d_C) is set to 5.0 Å, but can be defined by the user.

6.3.3 Topology analysis

In ProNA3D, the secondary structure of nucleic acids is represented as topology plots, which depict their global architecture as simplified two-dimensional graphs.

To generate topology plots, ProNA3D requires access to the executable of DSSR (Lu et al. 2015). DSSR is used to identify base pairs within the structure, including both canonical and non-canonical base-pairing. Upon execution of the topology plotting process, ProNA3D automatically creates a dedicated directory to store the DSSR output and executes DSSR. The resulting JSON files contain the secondary structure information required for topology plot generation. These files can be retained for subsequent downstream analyses beyond ProNA3D.

The parsing of DSSR-generated JSON files and the processing of information follow a workflow adapted from the RNAscape web server (Mitra et al. 2024), which also uses DSSR for determining the secondary structure in RNA and DNA. Once the JSON file has been generated and processed, ProNA3D generates the topology plot according to the selected base-pair annotation scheme. Four annotation modes are supported, corresponding to those provided by DSSR: the Leontis-Westhof classification (12 geometric classes, including canonical Watson-Crick base pairing as shown in Figure 1.10) (Leontis and Westhof 2001), the Saenger classification (28 classes) (Saenger 1988), the DSSR annotation (closely related to the Leontis-Westhof scheme with 12 classes) (Lu et al. 2015), and without base-pair annotations. RNA backbone connectivity is represented by solid black arrows connecting nucleotide nodes.

The resulting topology plots are fully interactive, allowing users to manually reposition nodes and to select residues or nucleotides in the corresponding three-dimensional structure directly from the two-dimensional plot. To facilitate the interpretation of secondary structure in the context of three-dimensional interfaces, ProNA3D enables coloring of topology plots according to sub-interface assignments. For complexes predicted by AlphaFold3, topology plots can be colored based on pLDDT values, allowing the assessment of prediction confidence directly within the two-dimensional representation.

For protein-nucleic acid interfaces, ProNA3D provides the option to incorporate protein contacts into the topology plot. Protein residues interacting with RNA or DNA are connected by edges to the corresponding nucleic acid nodes and can be directly selected from the topology plot in the associated three-dimensional structure. To enhance clarity and avoid overcrowding, each protein residue node is connected to only a single nucleic acid node. Consequently, protein residue nodes may be duplicated when necessary.

6.3.4 Two-dimensional interaction networks

In addition to topology plots, ProNA3D provides a simplified two-dimensional representation of the detected interface. In this representation, interacting residues and nucleotides are depicted as nodes, while their interactions are represented as edges. The visualization is implemented using the Python libraries `Matplotlib` (version 3.8.4) (Hunter 2007) and `NetworkX` (version 2.8.4) (Hagberg et al. 2008).

The resulting two-dimensional interface plot offers an overview of the interaction network and supports the analysis of connectivity, including at the level of individual

residues or nucleotides. The plot is fully interactive, enabling direct selection of node- and edge-associated entities in the three-dimensional structure.

6.3.5 AlphaFold3 scoring

ProNA3D supports the analysis of nucleic acid-containing complexes predicted by AlphaFold3. As in PICKLUSTER, users have to define a pLDDT threshold for interface detection (default: 50) (see subsection 6.1.1).

In addition, ProNA3D provides a set of scoring metrics for assessing model quality in AlphaFold3-predicted nucleic acid-containing complexes. These include ipLDDT (Equation 6.3), iPAE (Equation 6.4), pTM, ipTM, the *model confidence*, the maximum PAE value, and the full PAE matrix. Similar to PICKLUSTER, the calculation of these metrics, except for ipLDDT, requires access to the score-containing JSON files generated by AlphaFold3.

Visualization options for these metrics, including pLDDT-based coloring of interface residues and nucleotides in both the three-dimensional structure and the topology plot, are described in section 6.1.3 and subsection 6.3.3.

6.3.6 Cryo-EM map analysis

For cryo-EM-derived structures, ProNA3D supports the interface-specific analysis of density maps by restricting the map to regions surrounding the identified interfaces. This functionality relies on the volume zoning feature (ChimeraX command: *volume zone*) implemented in UCSF ChimeraX (Pettersen et al. 2021). When density zoning is enabled in ProNA3D, the displayed map surface is limited to a region within 5.0 Å of any atom belonging to the selected sub-interface. Cryo-EM density maps are obtained from the Electron Microscopy Data Bank (EMDB) and used as input together with the corresponding atomic models, enabling interface-specific zoning directly within the ProNA3D plug-in.

Local resolution was computed using RELION 4.0.0 (Kimanius et al. 2021) based on half-map reconstructions.

6.3.7 Dataset curation for large-scale PDB analysis

PDB entries were initially filtered for structures containing nucleic acids with a resolution of ≤ 3.0 Å and a release date prior to 04.02.2026, resulting in a set of 10,806 complexes. To reduce sequence redundancy, MMseqs2 (Steinegger and Söding 2017) was applied in a multi-step procedure. Individual chains were first categorized by macromolecular type (protein, DNA, or RNA). Protein sequences, extracted from SEQRES records, were clustered at 30% sequence identity, while DNA and RNA sequences were clustered at a threshold of 90% identity.

Subsequently, ProNA3D was applied with a default d_{HA} to identify all potential interfaces within each complex, without performing interface-level clustering. The resulting interfaces were mapped to sequence cluster interactions. A representative structure was randomly selected for each cluster combination, yielding 6,642 complexes. A further filtering step imposed minimum sequence length requirements of ≥ 30 residues for proteins and ≥ 6 nucleotides for nucleic acids, resulting in a dataset of 6,047 complexes.

Application of ProNA3D with default d_{HA} and d_C parameters to this dataset identified a total of 59,959 sub-interfaces.

These sub-interfaces were categorized by interaction type: RNA-RNA ($n = 4,797$), RNA-DNA ($n = 450$), DNA-DNA ($n = 4,590$), protein-DNA ($n = 11,791$), and protein-RNA ($n = 23,616$). In addition, 14,681 interfaces were observed between protein chains only, and 34 sub-interfaces involved hybrid chains containing both protein residues and nucleotides that could not be unambiguously assigned to a single category (e.g., AA-tRNA). As ProNA3D is specifically designed to analyze nucleic acid-containing interactions, these two classes were excluded, resulting in a final dataset comprising 45,244 sub-interfaces.

6.3.8 Viral subset

To investigate whether viral nucleic acid-containing complexes exhibit distinct interface properties, a viral subset was extracted from the large-scale dataset (see subsection 6.3.7). Source organism annotations for each PDB entry were retrieved and mapped to the NCBI taxonomy database. For each structure, the complete taxonomic lineage was traced from the annotated organism to the root of the NCBI taxonomy tree. Structures were classified as viral if the NCBI TaxID 10239 (Viruses) was present at any level of the lineage.

Application of this pipeline yielded a viral subset comprising 2,397 nucleic acid-containing sub-interfaces, including RNA-RNA ($n = 198$), RNA-DNA ($n = 30$), DNA-DNA ($n = 329$), protein-DNA ($n = 938$), and protein-RNA ($n = 902$) sub-interfaces.

6.3.9 B-factor normalization

To enable comparison of B-factors across structures with different absolute B-factor ranges, values were normalized to a range between 0 and 1 on a per-structure basis using min-max scaling. For each structure, the minimum ($B_{\min}$) and maximum ($B_{\max}$) B-factor values were extracted from the mmCIF file. The normalized mean B-factor of an interface was then calculated as:

$$B_{\text{interface,norm}} = \frac{B_{\text{interface}} - B_{\min}}{B_{\max} - B_{\min}} \quad (6.5)$$

where $B_{\text{interface}}$ represents the mean B-factor of all residues or nucleotides participating in the interface. The analysis was restricted to X-ray crystallographic structures.

6.3.10 Connectivity analysis

The connectivity of a residue or nucleotide was defined as the number of distinct residues or nucleotides on the opposing chain with which it forms at least one heavy-atom contact within d_{HA} . Connectivity networks were derived from the two-dimensional interface graph generated by ProNA3D (see subsection 6.3.4). The maximum connectivity of an interface was defined as the highest connectivity value observed among all residues or nucleotides participating in that interface.

6.3.11 Mann-Whitney U test

Statistical significance between independent groups was assessed using the Mann-Whitney U test, a nonparametric method that does not require normally distributed data. The Mann-Whitney U test can be applied to independent samples and evaluates whether the distributions of the two groups differ in their central tendency. A significance threshold of $\alpha = 0.05$ was used. For the pairwise comparisons of maximum connectivity distributions across interface classes, p-values were adjusted using the Benjamini-Hochberg false discovery rate correction (Benjamini and Hochberg 1995) to account for multiple testing. Values were computed using the stats module of SciPy (version 1.11.3) (Virtanen et al. 2020) and the `multipletests` function from the `statsmodels` library (version 0.14.0) (Seabold and Perktold 2010).

6.4 Data and software availability

All software and datasets generated as part of this thesis are publicly available as described below.

PICKLUSTER

PICKLUSTER is provided as a plug-in for UCSF ChimeraX (version 1.4 or later) and can be downloaded from the ChimeraX Toolshed. In addition, a command-line version is available for large-scale analyses. Both implementations are deposited in the project repository at <https://gitlab.com/topf-lab/pickluster.git>.

C2Qscore, individual prediction-based scores, ColabFold and AlphaFold3 predictions

C2Qscore and all individual prediction-based metrics are implemented as part of the PICKLUSTER v2.0 ChimeraX plug-in. A command-line version of C2Qscore is additionally available at <https://gitlab.com/topf-lab/c2qscore>. The repository also contains all ColabFold and AlphaFold3 predictions, along with the corresponding scoring metric values described in section 3.3.

ProNA3D

ProNA3D is implemented both as a ChimeraX (version 1.10, as well as earlier and later releases) plug-in and as a command-line tool. The ChimeraX plug-in can be installed via the ChimeraX Toolshed, while both versions are available in the project repository at <https://gitlab.com/topf-lab/ProNA3D>.

A

Additional Material

A.1 Supplemental figures

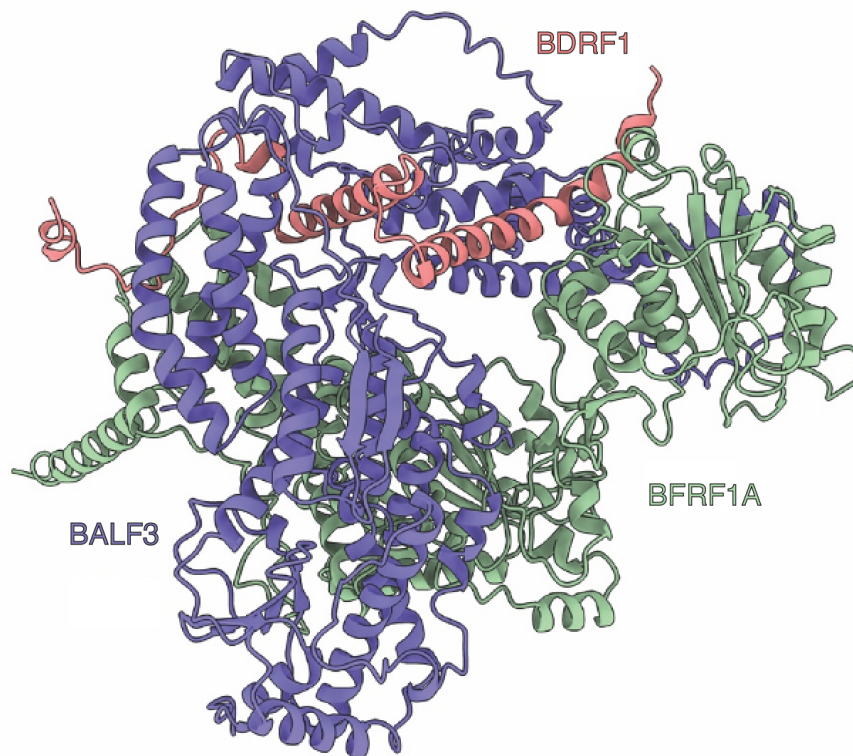


Figure A.1: BALF3 (purple), BDRF1 (salmon), and BFRF1A (green) from *Epstein-Barr Virus* modelled as a trimeric complex with ColabFold. Adapted from Genz et al. (2023).

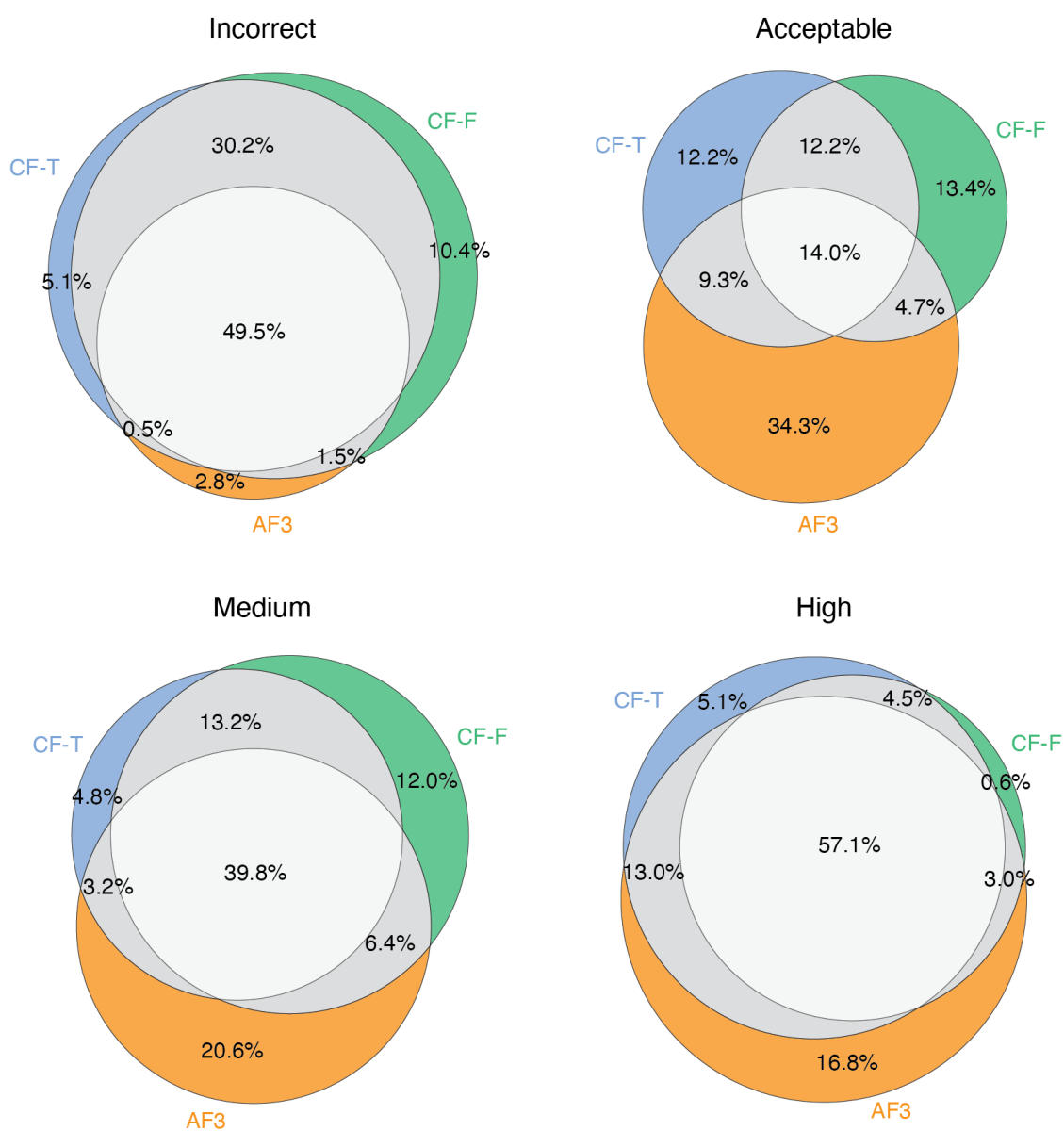


Figure A.2: Venn-diagrams of DockQ classifications for CF-T, CF-F and AF3. Adapted from Genz et al. (2025a).

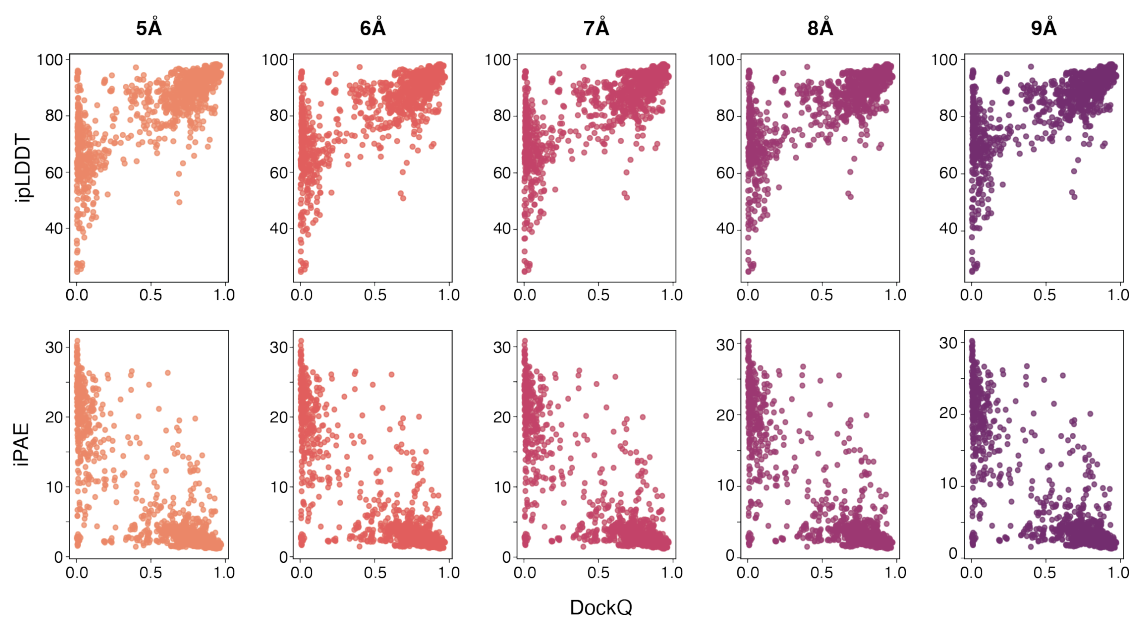


Figure A.3: Scatter plots of ipLDDT and iPAE versus DockQ across distance thresholds ranging from 5 to 9 Å. Adapted from Genz et al. (2025a).

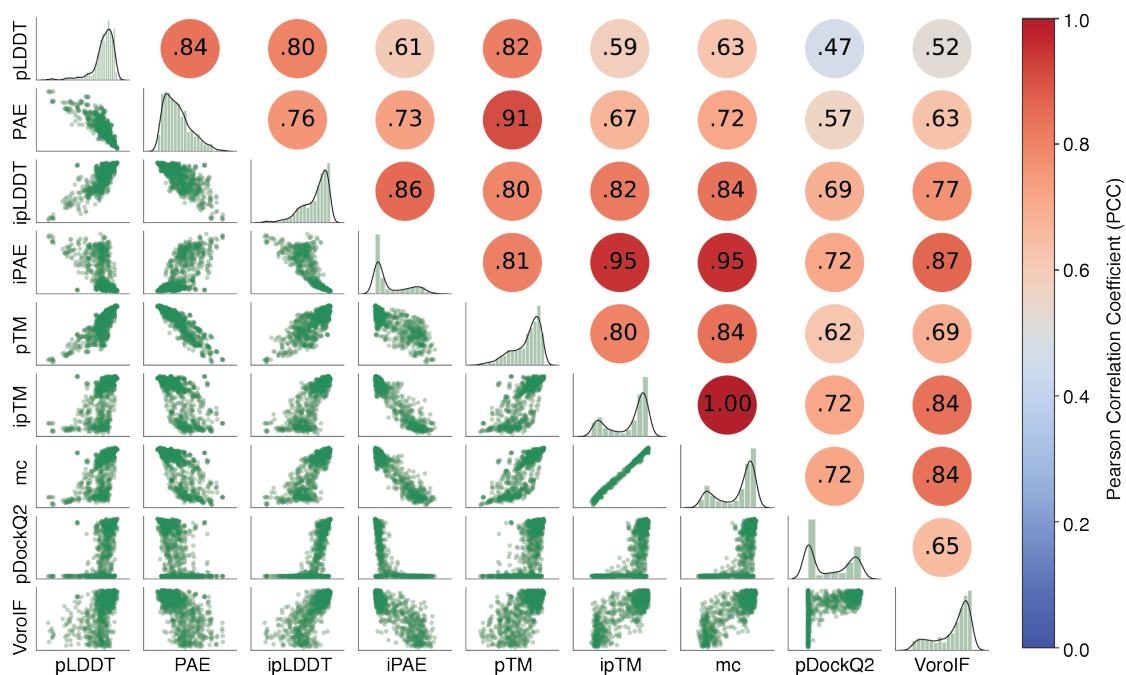


Figure A.4: Distributions and correlations of model quality scores for the CF-F predictions. The evaluated metrics include pLDDT, PAE, ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VoroIF. The diagonal panels show the distribution of each score across the benchmark set. Pairwise relationships between scores are shown in the lower-left panels. The upper-right panels report the corresponding Pearson correlation coefficients (PCCs), with red indicating stronger correlations and blue indicating weaker correlations. Adapted from Genz et al. (2025a).

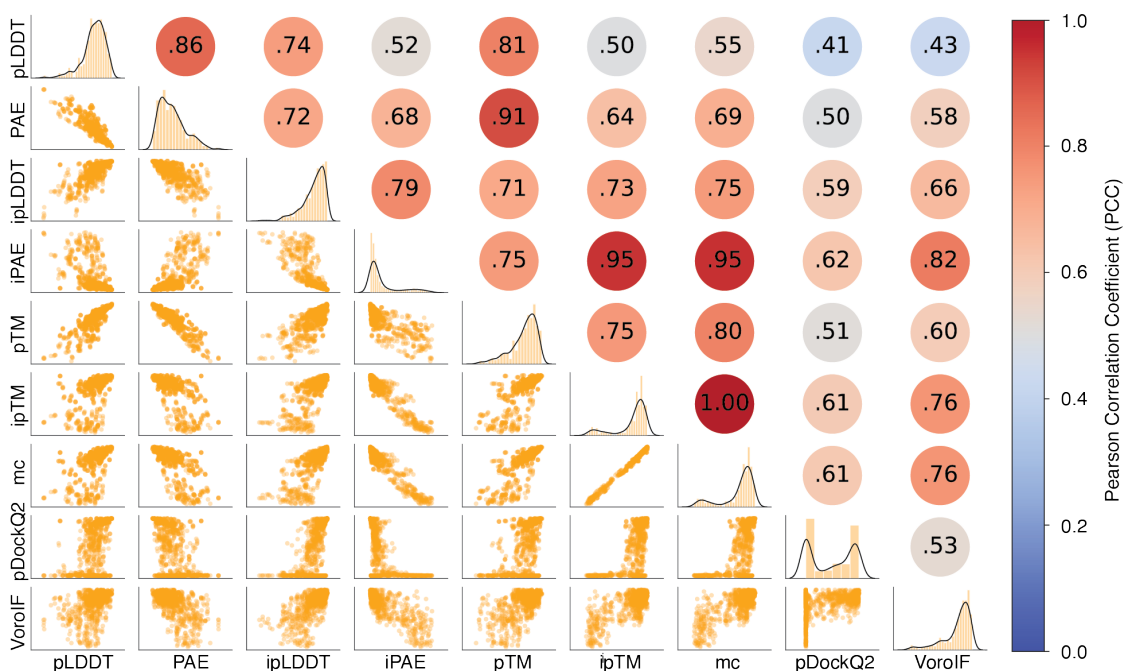


Figure A.5: Distributions and correlations of model quality scores for the AF3 predictions. The evaluated metrics include pLDDT, PAE, ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VorolF. The diagonal panels show the distribution of each score across the benchmark set. Pairwise relationships between scores are shown in the lower-left panels. The upper-right panels report the corresponding Pearson correlation coefficients (PCCs), with red indicating stronger correlations and blue indicating weaker correlations. Adapted from Genz et al. (2025a).

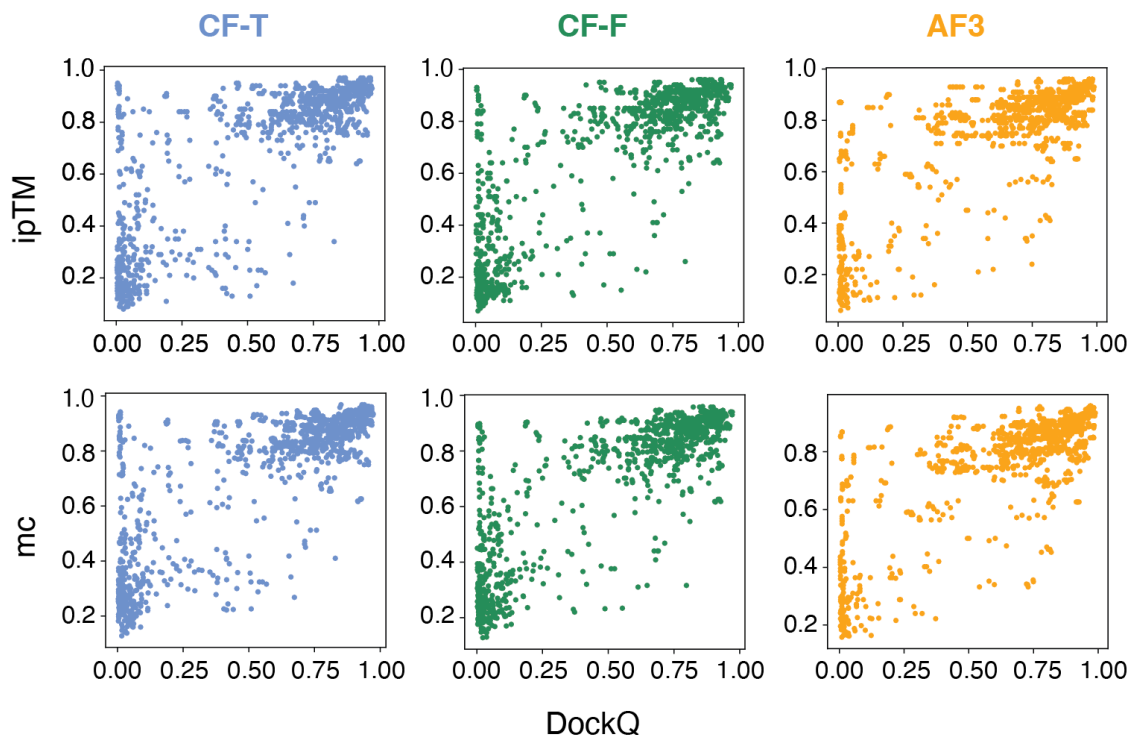


Figure A.6: ipTM and *model confidence* (mc) versus DockQ for the benchmark dataset across CF-T, CF-F, and AF3 predictions. Adapted from Genz et al. (2025a).

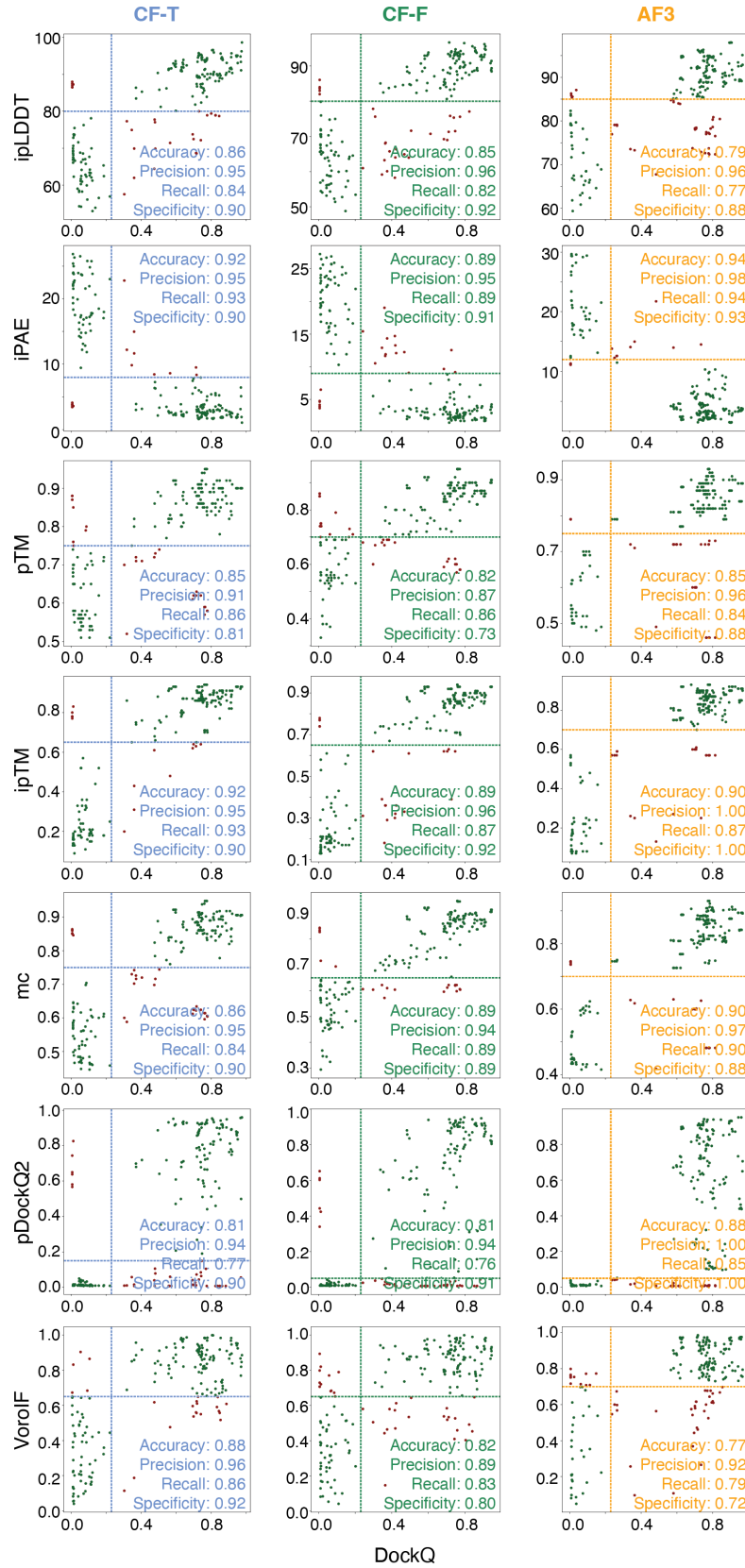


Figure A.7: Scatter plots of ipLDDT, iPAE, pTM, ipTM, *model confidence* (mc), pDockQ2, and VoroIF against DockQ for the test set across CF-T, CF-F, and AF3 predictions. True positives and true negatives are shown in dark green, whereas false positives and false negatives are shown in dark red. The cutoff values determined in subsection 3.3.4 are indicated by dotted lines. Adapted from Genz et al. (2025a).

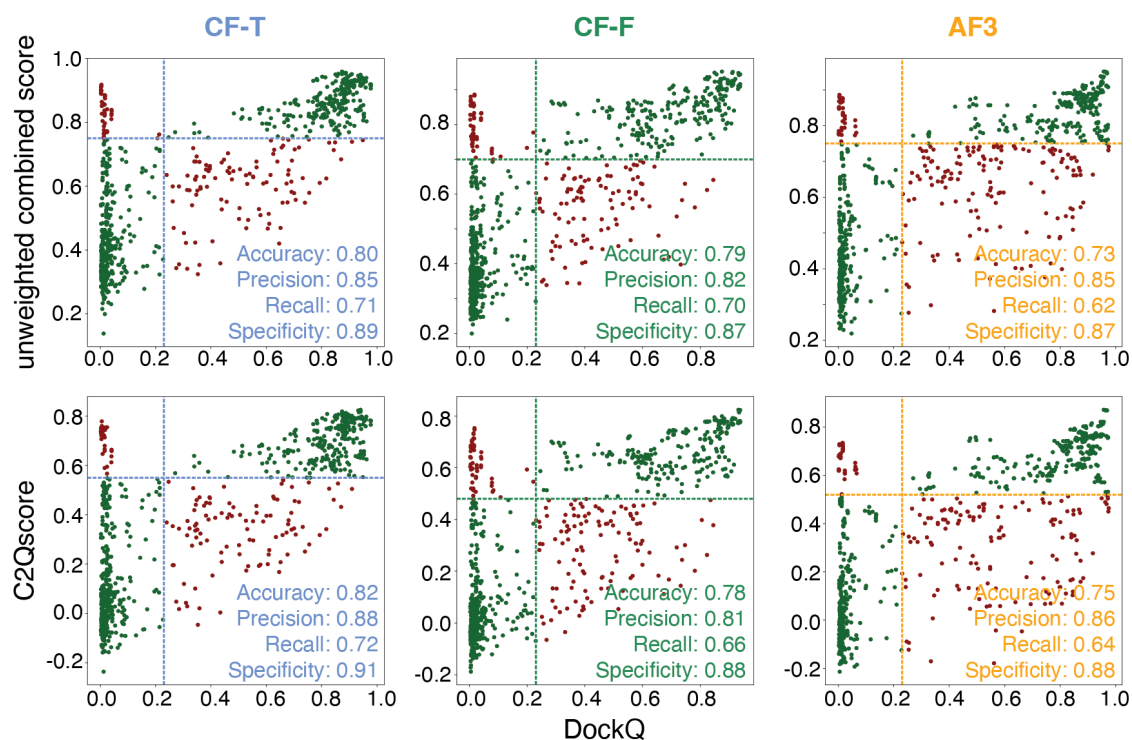


Figure A.8: Scatter plots of the unweighted combined score and C2Qscore for the large assembly test set across CF-T, CF-F, and AF3 predictions. True positives and true negatives are shown in dark green, whereas false positives and false negatives are shown in dark red. The determined cutoff values are indicated by dotted lines. C2Qscore values range from -0.45 to 0.88 for CF-T, from -0.37 to 0.87 for CF-F, and from -0.33 to 0.93 for AF3. Adapted from Genz et al. (2025a).

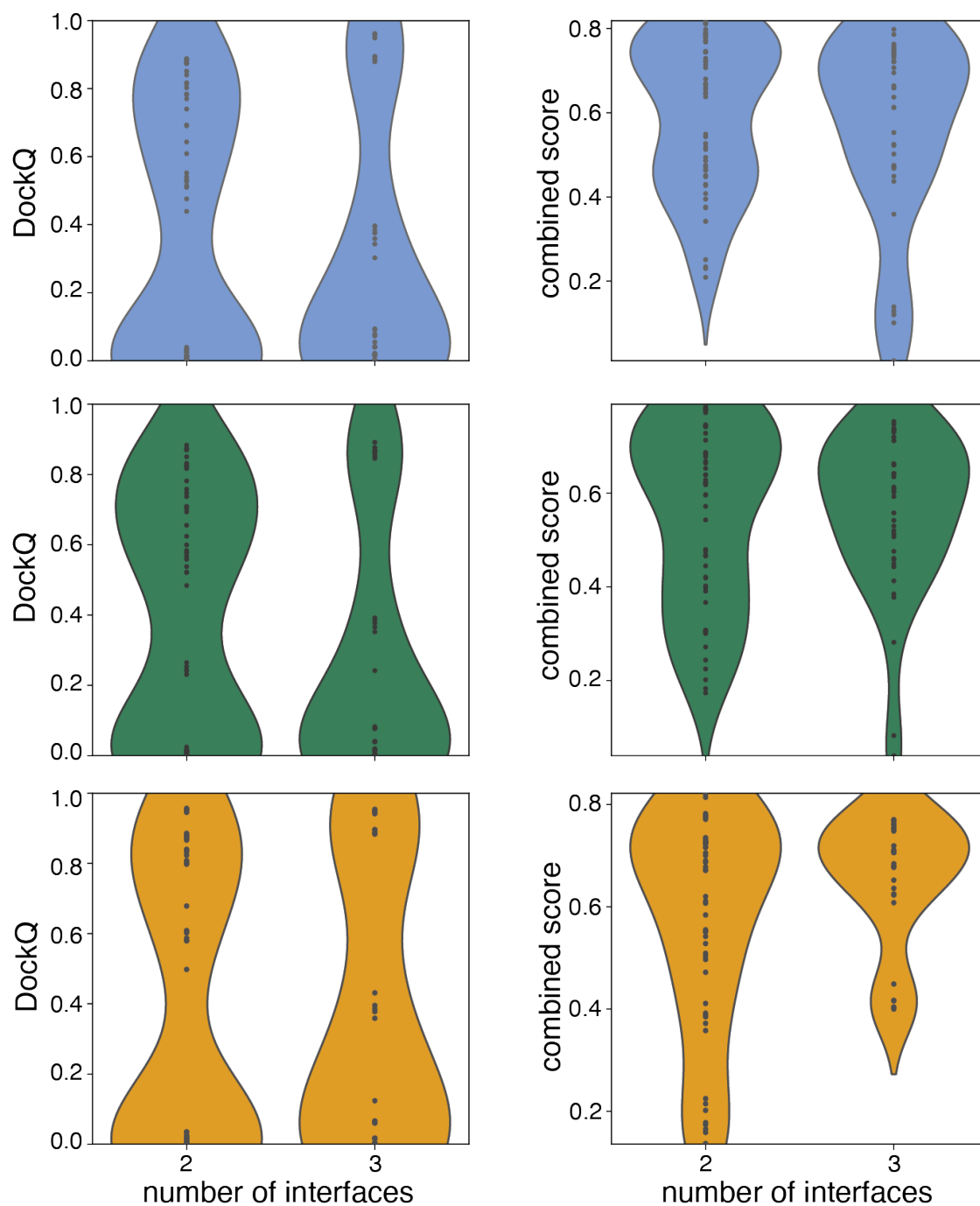


Figure A.9: Distribution of DockQ and C2Qscore values for CF-T (blue), CF-F (green), and AF3 (orange) predictions corresponding to cases with two (CF-T: $n = 58$, CF-F: $n = 57$, AF3: $n = 58$) or three (CF-T: $n = 42$, CF-F: $n = 39$, AF3: $n = 31$) potential configurations, for which the prediction resembles one of the alternative configurations. Adapted from Genz et al. (2025a).

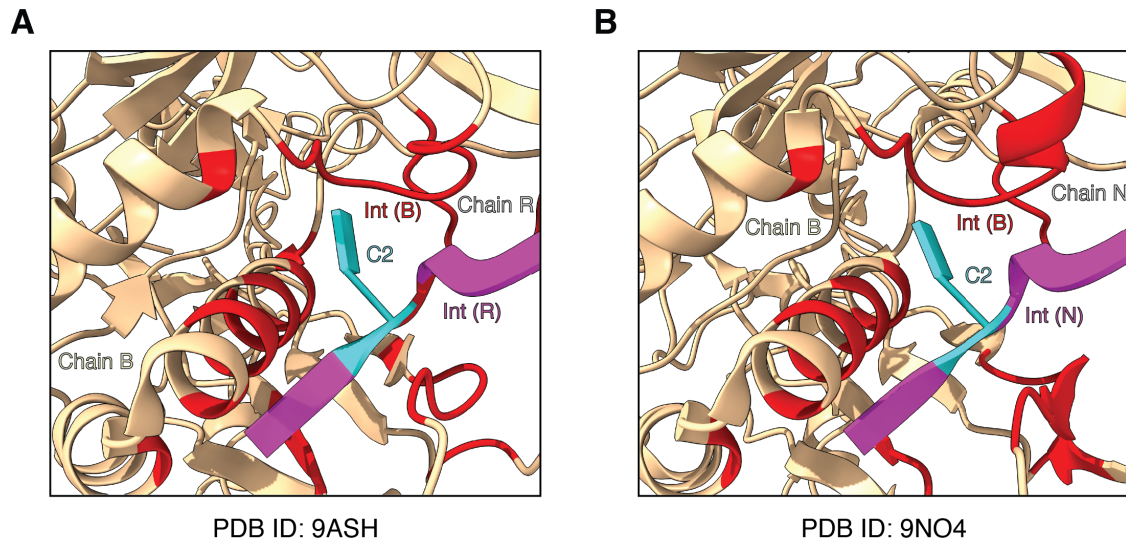


Figure A.10: Conserved nucleotide flipping associated with high-connectivity interactions in CRISPR-associated protein Csm4 complexes. Interface residues of Csm4 are shown in red, RNA interface nucleotides in magenta, and the flipped nucleotide with elevated connectivity in cyan. **(A)** Csm4 from *L. lactis* in complex with crRNA (PDB ID: 9ASH). **(B)** Csm4 from *S. thermophilus* in complex with crRNA (PDB ID: 9NO4). Adapted from Genz and Topf (2026).

A.2 Supplemental tables

Table A.1: Prediction-based scores used for the assessment of the structural models. Adapted from Genz et al. (2025a).

Score	Range	What is measured?	I/G
pLDDT (predicted local distance test) (Jumper et al. 2021)	0–100	per-residue measure of the local confidence; based on the local distance difference test	G
PAE (predicted alignment error) (Evans et al. 2022)	0–31.75 (capped) (Varadi et al. 2022)	expected positional error [Å] at residue X, if predicted and actual structures are aligned on residue Y (confidence of the relative positions)	G
ipLDDT (interface pLDDT) (Jumper et al. 2021)	0–100	mean pLDDT of interface residues (residues with heavy atoms in a distance of ≤ 5.0 Å from different chains) calculated using PICKLUSTER (Genz et al. 2023)	I
iPAE (interface PAE) (Evans et al. 2022)	0–31.75 (capped) (Varadi et al. 2022)	mean PAE of interface residues (residues with heavy atoms in a distance of ≤ 5.0 Å from different chains) calculated using PICKLUSTER (Genz et al. 2023)	I

continued on next page...

...continued from previous page

Score	Range	What is measured?	I/G
pTM (Jumper et al. 2021)	0–1	predicted TM-score (Y. Zhang and Skolnick 2004)	G
ipTM (Evans et al. 2022)	0–1	measures the accuracy of the predicted relative positions of the subunits forming the protein-protein complex	I
<i>model confidence</i> (Evans et al. 2022)	0–1	combination of pTM and ipTM: $0.8 \times \text{ipTM} + 0.2 \times \text{pTM}$	I+G
pDockQ2 (Zhu et al. 2023) https://gitlab.com/ElofssonLab/afm-benchmark	0–1	predicted interface DockQ; includes an interface-specific pLDDT and an interface-specific PAE	I
VoroIF-GNN (Olechnovič and Venclovas 2023) https://github.com/kliment-olechnovic/voronota	0–1	superposition-free score based on protein-protein interface graphs using Voronoi tessellation and a graph neural network; estimates the accuracy of protein assemblies	I

Table A.2: Accuracy and precision for individual scores for CF-T, CF-F and AF3. Binary classification: incorrect vs. (high, medium, acceptable). The used thresholds are reported in Figure 3.6. The calculations are based on a binary classification: incorrect vs. (high, medium, acceptable).

Score	CF-T		CF-F		AF3	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
ipLDDT	0.86	0.91	0.88	0.93	0.79	0.92
iPAE	0.90	0.93	0.91	0.94	0.91	0.94
pTM	0.83	0.89	0.85	0.87	0.81	0.95
ipTM	0.90	0.93	0.92	0.94	0.90	0.96
<i>model confidence</i>	0.88	0.95	0.92	0.94	0.90	0.96
pDockQ2	0.81	0.96	0.85	0.95	0.78	0.95
VoroIF	0.86	0.91	0.86	0.90	0.84	0.92
MEAN	0.86	0.93	0.88	0.92	0.85	0.94

Table A.3: Accuracy and precision for the gray zone of ipTM ($0.6 \leq \text{ipTM} \leq 0.8$) for CF-T, CF-F and AF3. The used threshold is reported in Figure 3.6A. The calculations are based on a binary classification: incorrect vs. (high, medium, acceptable).

Score	CF-T (n = 164)		CF-F (n = 180)		AF3 (n = 261)	
	Accuracy	Precision	Accuracy	Precision	Accuracy	Precision
$0.6 \leq \text{ipTM} \leq 0.8$	0.82	0.86	0.86	0.90	0.87	0.93

Table A.4: Pearson correlation of all approaches used to determine a combined score for CF-T, CF-F and AF3.

Approach	CF-T	CF-F	AF3
Unweighted – all scores	0.817	0.829	0.771
Unweighted – Pearson correlation-based	0.820	0.841	0.779
Linear regression – all scores	0.824	0.844	0.783
Linear regression – Pearson correlation-based	0.820	0.843	0.781
Lasso regression	0.815	0.837	0.776

Table A.5: Weights for C2Qscore for CF-T, CF-F and AF3.

Score	CF-T	CF-F	AF3
ipLDDT	0.561	0.446	-0.036
iPAE	0.123	0.082	0.169
pTM	-0.084	-0.084	0.335
ipTM	0.583	0.752	0.683
VoroIF	0.141	0.039	0.114

Table A.6: Bias for the unweighted combined score and C2Qscore for CF-T, CF-F and AF3.

Approach	CF-T	CF-F	AF3
Unweighted – all scores	–	–	–
C2Qscore (linear regression – all scores)	-0.448	-0.367	-0.331

Table A.7: Maximal values for the unweighted combined score and C2Qscore for CF-T, CF-F and AF3.

Approach	CF-T	CF-F	AF3
Unweighted – all scores	1.000	1.000	1.000
C2Qscore (linear regression – all scores)	0.878	0.867	0.933

Table A.8: P-values for the Wilcoxon signed-rank test of C2Qscore against the unweighted combination of all scores.

Comparison	CF-T	CF-F	AF3
Unweighted vs. C2Qscore	5.88×10^{-184}	5.90×10^{-184}	5.88×10^{-184}

Table A.9: Determined thresholds for the unweighted combined score and C2Qscore for distinguishing models of low (incorrect DockQ) and high (high, medium, acceptable DockQ) quality for CF-T, CF-F and AF3.

Approach	CF-T	CF-F	AF3
Unweighted – all scores	0.75	0.70	0.75
C2Qscore (linear regression – all scores)	0.55	0.48	0.52

Table A.10: Interface size statistics, sorted by interface type. For protein-nucleic acid interfaces, protein and nucleic acid sides are listed separately. n denotes the number of sub-interfaces; σ denotes the standard deviation.

Interface type	n	Mean	Median	σ
RNA-RNA (total nucleotides)	4,797	11.75	7.0	17.97
DNA-DNA (total nucleotides)	4,590	24.55	20.0	32.97
RNA-DNA (RNA side)	450	10.91	9.0	6.8
RNA-DNA (DNA side)	450	10.60	9.0	6.4
Protein-DNA (protein side)	11,791	11.11	8.0	11.3
Protein-DNA (DNA side)	11,791	4.69	4.0	3.4
Protein-RNA (protein side)	23,616	13.97	4.0	22.8
Protein-RNA (RNA side)	23,616	10.73	3.0	18.5

Table A.11: Normalized B-factor statistics, restricted to X-ray crystallographic structures, stratified by interface type. B-factors are normalized per structure using min-max scaling. n denotes the number of sub-interfaces; σ denotes the standard deviation.

Interface type	n	Mean	Median	σ
RNA-RNA	1,230	0.32	0.30	0.16
RNA-DNA	157	0.28	0.23	0.16
DNA-DNA	4,001	0.34	0.32	0.15
Protein-DNA	8,045	0.29	0.26	0.14
Protein-RNA	3,244	0.29	0.26	0.15

Table A.12: Maximum connectivity across all interface types. Shown are the number of sub-interfaces analyzed (n) and the distribution of maximal connectivity values, reported as median, mean, and standard deviation σ .

Interface type	n	Median	Mean	σ
RNA-RNA	4,797	3.00	3.22	1.83
RNA-DNA	450	3.00	3.03	0.76
DNA-DNA	4,590	3.00	2.92	0.88
Protein-DNA	11,791	5.00	5.52	3.48
Protein-RNA	23,616	4.00	4.64	3.32

Table A.13: Mann-Whitney U (Wilcoxon rank-sum) test for maximum connectivity interface type populations. P-values are Benjamini-Hochberg corrected for multiple testing ($n = 10$ pairwise comparisons).

P-value	RNA-RNA	DNA-DNA	RNA-DNA	Protein-DNA	Protein-RNA
RNA-RNA	–	1.31×10^{-1}	3.84×10^{-1}	$\ll 0.001$	1.70×10^{-109}
DNA-DNA	–	–	1.30×10^{-1}	$\ll 0.001$	4.67×10^{-114}
RNA-DNA	–	–	–	6.23×10^{-46}	2.61×10^{-10}
Protein-DNA	–	–	–	–	1.26×10^{-124}
Protein-RNA	–	–	–	–	–

Table A.14: Connectivity statistics by nucleotide type for DNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	28,518	3.00	2.82	0.60
T	28,153	3.00	2.78	0.61
G	28,473	3.00	2.75	0.60
C	27,562	3.00	2.68	0.61

Table A.15: Connectivity statistics by nucleotide type for RNA-RNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	15,874	2.00	2.44	1.31
U	12,110	2.00	2.31	1.17
G	15,056	2.00	2.34	1.17
C	13,342	2.00	2.13	1.10

Table A.16: Connectivity statistics by nucleotide type (RNA) for RNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	1,364	3.00	2.60	0.63
U	1,047	2.00	2.31	0.68
G	1,436	3.00	2.47	0.72
C	1,063	2.00	2.39	0.63

Table A.17: Connectivity statistics by nucleotide type (DNA) for RNA-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	1,014	3.00	2.64	0.66
T	1,351	3.00	2.49	0.70
G	1,095	3.00	2.64	0.62
C	1,309	2.00	2.39	0.66

Table A.18: Connectivity statistics by nucleotide type (DNA) for protein-DNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	14,097	3.00	3.88	2.78
T	15,072	4.00	4.15	2.81
G	13,374	4.00	4.20	2.96
C	12,805	3.00	3.84	2.74

Table A.19: Connectivity statistics by residue type (protein) for protein-DNA interfaces. n denotes the number of residues; σ denotes the standard deviation.

Residue	n	Median	Mean	σ
ASP	4,807	1.00	1.46	0.63
GLU	5,114	1.00	1.51	0.68
ALA	5,898	1.00	1.49	0.59
ARG	20,052	2.00	2.02	0.82
ASN	7,539	2.00	1.73	0.72
CYS	845	2.00	1.58	0.62
GLN	6,056	2.00	1.75	0.72
GLY	8,843	1.00	1.55	0.60
HIS	4,070	2.00	1.72	0.76
ILE	4,443	1.00	1.51	0.60
LEU	5,264	1.00	1.47	0.59
LYS	17,042	2.00	1.74	0.72
MET	2,123	2.00	1.66	0.68
PHE	4,061	2.00	1.63	0.63
PRO	4,337	1.00	1.51	0.59
SER	9,791	2.00	1.69	0.64
THR	7,872	2.00	1.70	0.65
TRP	1,945	2.00	1.74	0.73
TYR	6,092	2.00	1.86	0.80
VAL	4,789	1.00	1.53	0.60

Table A.20: Connectivity statistics by nucleotide type (RNA) for protein-RNA interfaces. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Median	Mean	σ
A	68,085	2.00	2.96	2.26
U	54,827	2.00	3.09	2.35
G	73,444	2.00	2.93	2.16
C	56,965	2.00	2.89	2.11

Table A.21: Connectivity statistics by residue type (protein) for protein-RNA interfaces. n denotes the number of residues; σ denotes the standard deviation.

Residue	n	Median	Mean	σ
ASP	8,397	1.00	1.75	1.04
GLU	9,224	1.00	1.76	1.01
ALA	15,824	2.00	1.95	1.12
ARG	53,411	2.00	2.87	1.62
ASN	16,357	2.00	2.21	1.21
CYS	2,806	2.00	1.87	1.15
GLN	12,961	2.00	2.27	1.32
GLY	23,561	2.00	2.08	1.07
HIS	12,987	2.00	2.58	1.48
ILE	11,211	2.00	1.86	1.09
LEU	15,944	2.00	1.85	1.01
LYS	49,536	2.00	2.53	1.41
MET	6,315	2.00	2.33	1.48
PHE	10,426	2.00	2.17	1.30
PRO	13,701	2.00	2.05	1.19
SER	18,946	2.00	2.12	1.14
THR	16,630	2.00	2.11	1.17
TRP	3,878	2.00	2.38	1.61
TYR	12,774	2.00	2.24	1.30
VAL	15,037	2.00	1.86	1.10

Table A.22: Interface size statistics for the viral subset, stratified by interface type. For protein-nucleic acid interfaces, protein and nucleic acid sides are listed separately. n denotes the number of sub-interfaces; σ denotes the standard deviation.

Interface type	n	Mean	Median	σ
RNA-RNA (total nucleotides)	198	16.32	13.0	14.34
DNA-DNA (total nucleotides)	329	25.78	26.0	20.91
RNA-DNA (RNA side)	30	12.43	10.5	7.1
RNA-DNA (DNA side)	30	12.40	10.5	6.7
Protein-DNA (protein side)	938	13.05	9.0	12.8
Protein-DNA (DNA side)	938	5.05	4.0	3.7
Protein-RNA (protein side)	902	15.56	7.0	21.7
Protein-RNA (RNA side)	902	7.51	4.0	14.1

Table A.23: Normalized B-factor statistics for the viral subset, restricted to X-ray crystallographic structures, stratified by interface type. B-factors are normalized per structure using min-max scaling. n denotes the number of sub-interfaces included in the B-factor analysis; σ denotes the standard deviation.

Interface type	n	Mean	Median	σ
RNA-RNA	73	0.41	0.40	0.19
RNA-DNA	11	0.53	0.52	0.16
DNA-DNA	284	0.38	0.35	0.17
Protein-DNA	694	0.32	0.29	0.16
Protein-RNA	260	0.29	0.26	0.13

Table A.24: Maximum connectivity statistics for the viral subset, stratified by interface type. Maximum connectivity is defined as the highest number of inter-partner contacts formed by any single residue or nucleotide at a given sub-interface. n denotes the number of sub-interfaces; σ denotes the standard deviation.

Interface type	n	Mean	Median	σ
RNA-RNA	198	3.15	3.00	1.38
RNA-DNA	30	3.10	3.00	0.65
DNA-DNA	329	3.00	3.00	0.71
Protein-RNA	902	5.80	5.00	4.13
Protein-DNA	938	5.76	6.00	3.46

Table A.25: Per-nucleotide mean connectivity at DNA-DNA interfaces in the viral subset. Connectivity is defined as the number of contacts formed by each nucleotide with its interface partner. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
DA	2,129	2.84	3.00	0.50
DT	2,165	2.77	3.00	0.61
DG	2,122	2.76	3.00	0.56
DC	2,067	2.69	3.00	0.60

Table A.26: Per-nucleotide mean connectivity at RNA-RNA interfaces in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
A	778	2.64	3.00	0.92
G	889	2.53	3.00	0.86
U	718	2.45	3.00	0.88
C	846	2.29	2.00	0.85

Table A.27: Per-nucleotide mean connectivity at RNA-DNA interfaces (RNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
G	103	2.75	3.00	0.53
A	90	2.56	3.00	0.63
C	94	2.49	3.00	0.58
U	86	2.34	2.00	0.69

Table A.28: Per-nucleotide mean connectivity at RNA-DNA interfaces (DNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
DG	99	2.74	3.00	0.56
DA	80	2.62	3.00	0.58
DC	101	2.50	3.00	0.64
DT	92	2.33	2.00	0.71

Table A.29: Per-nucleotide mean connectivity at protein-DNA interfaces (DNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
DT	1,392	4.50	4.00	2.84
DG	1,100	4.31	4.00	2.80
DA	1,185	4.10	4.00	2.74
DC	1,064	4.01	4.00	2.50

Table A.30: Per-nucleotide mean connectivity at protein-RNA interfaces (RNA side) in the viral subset. n denotes the number of nucleotides; σ denotes the standard deviation.

Nucleotide	n	Mean	Median	σ
U	1,513	4.32	4.00	3.13
A	1,738	3.97	3.00	3.04
C	1,624	3.66	3.00	2.83
G	1,898	3.56	3.00	2.84

Table A.31: Per-residue mean connectivity at protein-DNA interfaces (protein side) in the viral subset. n denotes the number of residues; σ denotes the standard deviation.

Residue	n	Median	Mean	σ
ALA	582	1.00	1.47	0.53
ASP	446	1.00	1.39	0.56
GLU	455	1.00	1.49	0.61
GLY	961	1.00	1.47	0.55
ILE	424	1.00	1.35	0.51
LEU	407	1.00	1.36	0.57
PRO	564	1.00	1.47	0.60
TRP	156	1.00	1.54	0.67
VAL	444	1.00	1.53	0.58
ARG	1,373	2.00	1.93	0.78
ASN	773	2.00	1.67	0.73
CYS	116	2.00	1.77	0.44
GLN	623	2.00	1.61	0.70
HIS	391	2.00	1.70	0.67
LYS	1,880	2.00	1.78	0.71
MET	189	2.00	1.70	0.72
PHE	386	2.00	1.67	0.60
SER	810	2.00	1.69	0.62
THR	584	2.00	1.65	0.65
TYR	675	2.00	1.81	0.74

Table A.32: Per-residue mean connectivity at protein-RNA interfaces (protein side) in the viral subset. n denotes the number of residues; σ denotes the standard deviation.

Residue	n	Median	Mean	σ
ALA	645	1.00	1.65	0.86
ASP	636	1.00	1.48	0.71
GLU	469	1.00	1.45	0.65
THR	811	1.00	1.65	0.86
TRP	166	1.00	1.75	1.19
VAL	563	1.00	1.62	0.86
ARG	1,805	2.00	2.39	1.41
ASN	841	2.00	1.84	0.93
CYS	170	2.00	1.63	0.73
GLN	532	2.00	1.80	0.95
GLY	930	2.00	1.75	0.83
HIS	430	2.00	2.14	1.31
ILE	489	2.00	1.71	0.89
LEU	692	2.00	1.66	0.83
LYS	1,807	2.00	2.08	1.18
MET	363	2.00	1.85	1.00
PHE	413	2.00	1.87	1.06
PRO	501	2.00	1.74	0.93
SER	1,119	2.00	1.76	0.88
TYR	657	2.00	1.88	0.98

A.3 Supplemental text

Reasons for score calculation failures. A small number of models were excluded from the large assembly set due to score calculation failures caused by segmentation faults, invalid .pdb coordinate entries, or the absence of detectable protein-protein interfaces by PICKLUSTER (subsection 6.1.2). This resulted in minor differences in model counts across the prediction sets (CF-T: 768, CF-F: 769, AF3: 720).

B

Software and Tools

The following overview lists all software, tools, libraries, and databases used in this dissertation, organized by functional category.

Name	Reference/Source
<i>Structure prediction</i>	
AlphaFold2	(Jumper et al. 2021; Evans et al. 2022).
AlphaFold3	(Abramson et al. 2024).
ColabFold	(Mirdita et al. 2022).
<i>Interface analysis</i>	
PICKLUSTER	(Genz et al. 2023).
ProNA3D	(Genz and Topf 2026).
UCSF ChimeraX	(Pettersen et al. 2021).
<i>Quality assessment</i>	
C2Qscore	(Genz et al. 2025a).
DockQ	(Basu and Wallner 2016; Mirabello and Wallner 2024).
pDockQ2	(Zhu et al. 2023).
VoroIF-GNN	(Olechnovič and Venclovas 2023).
<i>Sequence analysis</i>	
MMseqs2	(Steinegger and Söding 2017).
CD-HIT	(Fu et al. 2012).
<i>RNA/DNA structural analysis</i>	
DSSR	(Lu et al. 2015).
RNAscape	(Mitra et al. 2024).

continued on next page...

...continued from previous page

Name	Reference
<i>Cryo-EM processing</i>	
RELION	(Kimanius et al. 2021).
<i>Python libraries</i>	
SciPy	(Virtanen et al. 2020).
scikit-learn	(Pedregosa et al. 2011).
Gemmi	(Wojdyr 2022).
Matplotlib	(Hunter 2007).
NetworkX	(Hagberg et al. 2008).
statsmodels	(Seabold and Perktold 2010).
<i>Databases</i>	
Protein Data Bank (PDB)	(Berman et al. 2000).
Electron Microscopy Data Bank (EMDB)	(wwPDB Consortium 2024).
SABDab	(Dunbar et al. 2014).
NCBI Taxonomy Database	(Federhen 2012).
<i>Language and writing assistance*</i>	
Claude	Anthropic.
ChatGPT 5.5	OpenAI.
Perplexity AI	Perplexity AI, Inc.
Grammarly	Grammarly Inc.
<i>Figure preparation</i>	
Adobe Illustrator 2026	Adobe Inc.
ChemDraw 21.0.0	PerkinElmer Informatics.

* All tools listed under *Writing and grammar assistance* were used to a very limited extent to improve the phrasing of individual sentences.

C

Hazardous Substances

No chemicals were used.

References

- Abramson, J. et al. (2024). “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* 630, pp. 493–500. DOI: 10.1038/s41586-024-07487-w.
- Adasme, M. F., K. L. Linnemann, S. N. Bolz, F. Kaiser, S. Salentin, V. J. Haupt, and M. Schroeder (2021). “PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA”. In: *Nucleic Acids Res.* 49.W1, W530–W534. DOI: 10.1093/nar/gkab294.
- Adhav, V. A. and K. Saikrishnan (2023). “The Realm of Unconventional Noncovalent Interactions in Proteins: Their Significance in Structure and Function”. In: *ACS Omega* 8.25, pp. 22268–22284. DOI: 10.1021/acsomega.3c00205.
- Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams (2012). “Towards automated crystallographic structure refinement with phenix.refine”. In: *Acta Crystallogr. D Biol. Crystallogr.* 68.Pt 4, pp. 352–367. DOI: 10.1107/S0907444912001308.
- Akuh, O.-A., D. Maus, M. Blume, K. J. Saliba, and F. Seeber (2025). “Reconstituted ferredoxin-MEP pathway of Apicomplexa in *E. coli* as an in situ screening platform for inhibitors and essential enzyme mutations”. In: *J. Biol. Chem.* 301.11, 110726. DOI: 10.1016/j.jbc.2025.110726.
- Akwelle, R. N. et al. (2025). “In silico design and computational characterization of novel chimeric multiepitope antigens for Mpox serosurveillance”. In: *bioRxiv*.
- Alberts, B. (2003). “DNA replication and recombination”. In: *Nature* 421.6921, pp. 431–435. DOI: 10.1038/nature01407.
- Alonso, D. and A. Mondragón (2021). “Mechanisms of catalytic RNA molecules”. In: *Biochem. Soc. Trans.* 49.4, pp. 1529–1535. DOI: 10.1042/BST20200465.
- Álvarez-Salmoral, D., R. Borza, C. Maiella, B. P. Y. Kwee, R. Xie, R. P. Joosten, M. L. Hekkelman, and A. Perrakis (2024). “AlphaBridge: tools for the analysis of predicted biomolecular complexes”. In: *bioRxiv*. DOI: 10.1101/2024.10.23.619601.
- Anashkina, A. A. (2023). “Protein-DNA recognition mechanisms and specificity”. In: *Biophys. Rev.* 15.5, pp. 1007–1014. DOI: 10.1007/s12551-023-01137-7.
- Arakawa, T., T. Oyama, T. Ura, S. Nishinami, K. Shiraki, and T. Akuta (2025). “Differences and similarities in protein and nucleic acid structures and their biological interactions”. In: *Curr. Issues Mol. Biol.* 47.12, 1019. DOI: 10.3390/cimb47121019.
- Arnez, J. G. and J. Cavarelli (1997). “Structures of RNA-binding proteins”. In: *Q. Rev. Biophys.* 30.3, pp. 195–240. DOI: 10.1017/s0033583597003351.
- Aung, Z., S.-H. Tan, S.-K. Ng, and K.-L. Tan (2008). “PPiClust: EFFICIENT CLUSTERING OF 3D PROTEIN-PROTEIN INTERACTION INTERFACES”. In: *J. Bioinform. Comput. Biol.* 06.03, pp. 415–433. DOI: 10.1142/S0219720008003485.
- Aymami, J., M. Coll, G. A. van der Marel, J. H. van Boom, A. H. Wang, and A. Rich (1990). “Molecular structure of nicked DNA: a substrate for DNA repair enzymes”. In: *Proc. Natl. Acad. Sci. U. S. A.* 87.7, pp. 2526–2530. DOI: 10.1073/pnas.87.7.2526.

- Baek, M., I. Anishchenko, I. R. Humphreys, Q. Cong, D. Baker, and F. DiMaio (2023a). “Efficient and accurate prediction of protein structure using RoseTTAFold2”. In: *bioRxiv*. DOI: 10.1101/2023.05.24.542179.
- Baek, M., R. McHugh, I. Anishchenko, H. Jiang, D. Baker, and F. DiMaio (2023b). “Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA”. In: *Nat. Methods* 21, pp. 117–121. DOI: 10.1038/s41592-023-02086-5.
- Baek, M. et al. (2021). “Accurate prediction of protein structures and interactions using a three-track neural network”. In: *Science* 373.6557, pp. 871–876. DOI: 10.1126/science.abj8754.
- Bai, X.-C., I. S. Fernandez, G. McMullan, and S. H. W. Scheres (2013). “Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles”. In: *Elife* 2, e00461. DOI: 10.7554/eLife.00461.
- Ban, X., P. Lahiri, A. S. Dhoble, D. Li, Z. Gu, C. Li, L. Cheng, Y. Hong, Z. Li, and B. Kaustubh (2019). “Evolutionary stability of salt bridges hints its contribution to stability of proteins”. In: *Comput. Struct. Biotechnol. J.* 17, pp. 895–903. DOI: 10.1016/j.csbj.2019.06.022.
- Banco, M. T., T. Paul, J. Jiang, S. Myong, and A. R. Ferré-D’Amaré (2025). “Structural basis for dual DNA and RNA specificity of the G-quadruplex-resolving DEAH-box helicase DHX36”. In: *Cell Rep.* 44.9, 116136. DOI: 10.1016/j.celrep.2025.116136.
- Bansal, A., S. Kaushik, and S. Kukreti (2022). “Non-canonical DNA structures: Diversity and disease association”. In: *Front. Genet.* 13, 959258. DOI: 10.3389/fgene.2022.959258.
- Barhoon, M. and H. Mahdiuni (2025). “Exploring protein-protein docking tools: Comprehensive insights into traditional and deep-learning approaches”. In: *J. Chem. Inf. Model.* 65.13, pp. 6446–6469. DOI: 10.1021/acs.jcim.5c01029.
- Basu, S. and B. Wallner (2016). “DockQ: A Quality Measure for Protein-Protein Docking Models”. In: *PLoS One* 11.8, e0161879. DOI: 10.1371/journal.pone.0161879.
- Beard, P. M., N. S. Taus, and J. D. Baines (2002). “DNA cleavage and packaging proteins encoded by genes U(L)28, U(L)15, and U(L)33 of herpes simplex virus type 1 form a complex in infected cells”. In: *J. Virol.* 76.10, pp. 4785–4791. DOI: 10.1128/jvi.76.10.4785-4791.2002.
- Beckers, M., D. Mann, and C. Sachse (2021). “Structural interpretation of cryo-EM image reconstructions”. In: *Prog. Biophys. Mol. Biol.* 160, pp. 26–36. DOI: 10.1016/j.pbiomolbio.2020.07.004.
- Belios, I. et al. (2025). “Maternal *IGHG* locus duplications impair infants’ passive immunity”. In: *medRxiv*. DOI: 10.1101/2025.10.09.25337669.
- Bellone, M. L., F. Mensitieri, E. Marmo, A. N. Calabrese, G. Gaudino, V. Izzo, and F. Dal Piaz (2025). “Structural characterization of protein-nucleic acid complexes: An overview of the recent innovation in the analytic methods”. In: *Int. J. Mol. Sci.* 26.23, 11465. DOI: 10.3390/ijms262311465.
- Bender, B. J., A. Cisneros 3rd, A. M. Duran, J. A. Finn, D. Fu, A. D. Lokits, B. K. Mueller, A. K. Sangha, M. F. Sauer, A. M. Sevy, G. Sliwoski, J. H. Sheehan, F. DiMaio, J. Meiler, and R. Moretti (2016). “Protocols for molecular modeling with Rosetta3 and RosettaScripts”. In: *Biochemistry* 55.34, pp. 4748–4763. DOI: 10.1021/acs.biochem.6b00444.
- Benjamini, Y. and Y. Hochberg (1995). “Controlling the false discovery rate: A practical and powerful approach to multiple testing”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 57.1, pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

- Bergendahl, L. T., L. Gerasimavicius, J. Miles, L. Macdonald, J. N. Wells, J. P. I. Welburn, and J. A. Marsh (2019). "The role of protein complexes in human genetic disease: Protein Complexes in Genetic Disease". In: *Protein Sci.* 28.8, pp. 1400–1411. DOI: 10.1002/pro.3667.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne (2000). "The Protein Data Bank". In: *Nucleic Acids Res.* 29.1, pp. 235–242. DOI: 10.1093/nar/28.1.235.
- Beton, J. G., T. Mulvaney, T. Cragolini, and M. Topf (2024). "Cryo-EM structure and B-factor refinement with ensemble representation". In: *Nat. Commun.* 15, 444. DOI: 10.1038/s41467-023-44593-1.
- Biasini, M., T. Schmidt, S. Bienert, V. Mariani, G. Studer, J. Haas, N. Johner, A. D. Schenk, A. Philippsen, and T. Schwede (2013). "OpenStructure: an integrated software framework for computational structural biology". In: *Acta Crystallogr. D Biol. Crystallogr.* 69.Pt 5, pp. 701–709. DOI: 10.1107/S0907444913007051.
- Blair, H. A. (2021). "Sotorasib: First approval". In: *Drugs* 81.13, pp. 1573–1579. DOI: 10.1007/s40265-021-01574-2.
- Blundell, T. L., D. F. Burke, D. Chirgadze, V. Dhanaraj, M. Hyvönen, C. A. Innis, E. Parisini, L. Pellegrini, M. Sayed, and B. L. Sibanda (2000). "Protein-protein interactions in receptor activation and intracellular signalling". In: *Biol. Chem.* 381, pp. 955–959. DOI: 10.1515/BC.2000.117.
- Blundell, T. L. (2021). "The first resolution revolution in protein structure analysis: X-ray diffraction of polypeptide conformations and globular protein folds in 1950s and 1960s". In: *Prog. Biophys. Mol. Biol.* 167, pp. 32–40. DOI: 10.1016/j.pbiomolbio.2021.09.002.
- Bogan, A. A. and K. S. Thorn (1998). "Anatomy of hot spots in protein interfaces". In: *J. Mol. Biol.* 280.1, pp. 1–9. DOI: 10.1006/jmbi.1998.1843.
- Bonilla, S. L. and J. S. Kieft (2022). "The promise of cryo-EM to explore RNA structural dynamics". In: *J. Mol. Biol.* 434.18, 167802. DOI: 10.1016/j.jmb.2022.167802.
- Bonvin, A. M. J. J. (2006). "Flexible protein-protein docking". In: *Curr. Opin. Struct. Biol.* 16.2, pp. 194–200. DOI: 10.1016/j.sbi.2006.02.002.
- Bou-Nader, C., K. M. Juma, A. Bothra, A. J. Brasington, R. Ghirlando, M. Suzuki, D. N. Garboczi, S. H. Leppla, and J. Zhang (2025). "Specificity and mechanism of the double-stranded RNA-specific J2 monoclonal antibody". In: *bioRxiv*. DOI: 10.1101/2025.05.09.649859.
- Branden, C.-I. and J. Tooze (1999). *Introduction to Protein Structure*. 2nd ed. Boca Raton, FL: CRC Press. ISBN: 9780815323051.
- Bret, H., J. Gao, D. J. Zea, J. Andreani, and R. Guérois (2024). "From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2". In: *Nat. Commun.* 15, 597. DOI: 10.1038/s41467-023-44288-7.
- Brown, M. D. and D. B. Sacks (2009). "Protein scaffolds in MAP kinase signalling". In: *Cell. Signal.* 21.4, pp. 462–469. DOI: 10.1016/j.cellsig.2008.11.013.
- Bryant, P., G. Pozzati, and A. Elofsson (2022). "Improved prediction of protein-protein interactions using AlphaFold2". In: *Nat. Commun.* 13, 1265. DOI: 10.1038/s41467-022-28865-w.
- Bu, F. et al. (2025). "RNA-Puzzles Round V: blind predictions of 23 RNA structures". In: *Nat. Methods* 22.2, pp. 399–411. DOI: 10.1038/s41592-024-02543-9.

- Burke, D. F. et al. (2023). "Towards a structurally resolved human protein interaction network". In: *Nat. Struct. Mol. Biol.* 30.2, pp. 216–225. doi: 10.1038/s41594-022-00910-8.
- Burley, S. K. (2000). "An overview of structural genomics". In: *Nat. Struct. Biol.* 7 Suppl.11, pp. 932–934. doi: 10.1038/80697.
- Burley, S. K., H. M. Berman, G. J. Kleywegt, J. L. Markley, H. Nakamura, and S. Velankar (2017). "Protein Data Bank (PDB): The single global macromolecular structure archive". In: *Methods Mol. Biol.* Methods in molecular biology (Clifton, N.J.) 1607, pp. 627–641. doi: 10.1007/978-1-4939-7000-1_26.
- Butcher, S. E. and A. M. Pyle (2011). "The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks". In: *Acc. Chem. Res.* 44.12, pp. 1302–1311. doi: 10.1021/ar200098t.
- Cardone, G., J. B. Heymann, and A. C. Steven (2013). "One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions". In: *J. Struct. Biol.* 184.2, pp. 226–236. doi: 10.1016/j.jusb.2013.08.002.
- Černý, J. and P. Hobza (2007). "Non-covalent interactions in biomacromolecules". In: *Phys. Chem. Chem. Phys.* doi: 10.1039/B704781A.
- Chakrabarti, P. and J. Janin (2002). "Dissecting protein-protein recognition sites". In: *Proteins* 47.3, pp. 334–343. doi: 10.1002/prot.10085.
- Chawla, M., U. Kalra, A. Petta, S. Sharma, A. R. Shaikh, L. Cavallo, and R. Oliva (2025). "COCOMAPS 2.0: a web server for identifying, analyzing, and visualizing atomic interactions at the interface of biomolecular complexes". In: *Bioinformatics* 41.12, btaf606. doi: 10.1093/bioinformatics/btaf606.
- Cheatham, T. E. and P. A. Kollman (1997). "Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes". In: *J. Am. Chem. Soc.* 119.21, pp. 4805–4825. doi: 10.1021/ja963641w.
- Chen, L., Q. Li, K. F. A. Nasif, Y. Xie, B. Deng, S. Niu, S. Pouriyeh, Z. Dai, J. Chen, and C. Y. Xie (2024). "AI-driven deep learning techniques in protein Structure Prediction". In: *Int. J. Mol. Sci.* 25.15, 8426. doi: 10.3390/ijms25158426.
- Chen, R., L. Li, and Z. Weng (2003). "ZDOCK: an initial-stage protein-docking algorithm". In: *Proteins* 52.1, pp. 80–87. doi: 10.1002/prot.10389.
- Cheng, F. et al. (2021). "Comprehensive characterization of protein-protein interactions perturbed by disease mutations". In: *Nat. Genet.* 53.3, pp. 342–353. doi: 10.1038/s41588-020-00774-y.
- Cheng, Y. (2015). "Single-particle cryo-EM at crystallographic resolution". In: *Cell* 161.3, pp. 450–457. doi: 10.1016/j.cell.2015.03.049.
- Cheng, Y., N. Grigorieff, P. A. Penczek, and T. Walz (2015). "A primer to single-particle cryo-electron microscopy". In: *Cell* 161.3, pp. 438–449. doi: 10.1016/j.cell.2015.03.050.
- Cherstvy, A. G. (2009). "Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups". In: *J. Phys. Chem. B* 113.13, pp. 4242–4247. doi: 10.1021/jp810009s.
- Chothia, C. and J. Janin (1975). "Principles of protein-protein recognition". In: *Nature* 256.5520, pp. 705–708. doi: 10.1038/256705a0.
- Clackson, T. and J. A. Wells (1995). "A hot spot of binding energy in a hormone-receptor interface". In: *Science* 267.5196, pp. 383–386. doi: 10.1126/science.7529940.

- Comeau, S. R., D. W. Gatchell, S. Vajda, and C. J. Camacho (2004). "ClusPro: a fully automated algorithm for protein-protein docking". In: *Nucleic Acids Res.* 32.Issue suppl_2, W96–9. DOI: 10.1093/nar/gkh354.
- Cozzetto, D., A. Kryshchuk, M. Ceriani, and A. Tramontano (2007). "Assessment of predictions in the model quality assessment category". In: *Proteins* 69.S8, pp. 175–183. DOI: 10.1002/prot.21669.
- Crick, F. (1970). "Central dogma of molecular biology". In: *Nature* 227.5258, pp. 561–563. DOI: 10.1038/227561a0.
- Cruz, J. A. et al. (2012). "RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction". In: *RNA* 18.4, pp. 610–625. DOI: 10.1261/rna.031054.111.
- Cui, H., Y. Wang, Y. Fu, X. Yu, W. Li, F. Lin, and W. Han (2025). "Precision in predicting protein-nucleic acid complexes: Establishing a benchmark data set and comparative metrics". In: *J. Chem. Inf. Model.* 65.18, pp. 9654–9671. DOI: 10.1021/acs.jcim.5c01372.
- Cunningham, B. C. and J. A. Wells (1991). "Rational design of receptor-specific variants of human growth hormone". In: *Proc. Natl. Acad. Sci. U. S. A.* 88.8, pp. 3407–3411. DOI: 10.1073/pnas.88.8.3407.
- D'Imprima, E. and W. Kühlbrandt (2021). "Current limitations to high-resolution structure determination by single-particle cryoEM". In: *Q. Rev. Biophys.* 54, e4. DOI: 10.1017/S0033583521000020.
- Dan-Dukor, G., S. Shang, G. O. Leighton, C. R. Travis, T. D. Schwochert, P. Agrawal, O. Ajasa, T. Li, M. L. Waters, G. D. Ginder, and D. C. Williams Jr (2025). "The role of multivalency in the association of the eight twenty-one protein 2 (ETO2) with the nucleosome remodeling and deacetylase (NuRD) complex". In: *Nucleic Acids Res.* 53.10, gkaf439. DOI: 10.1093/nar/gkaf439.
- Dapkūnas, J., A. Timinskas, K. Olechnovič, M. Margelevičius, R. Dičiūnas, and Č. Venclovas (2016). "The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures". In: *Bioinformatics* 33.6, pp. 935–937. DOI: 10.1093/bioinformatics/btw756.
- Dapkūnas, J., A. Timinskas, K. Olechnovič, M. Tomkuvienė, and Č. Venclovas (2024). "PPI3D: a web server for searching, analyzing and modeling protein-protein, protein-peptide and protein-nucleic acid interactions". In: *Nucleic Acids Res.* 52.W1, W264–W271. DOI: 10.1093/nar/gkae278.
- Dapkūnas, J. and Č. Venclovas (2025). "Databases and web-based tools for studying structures of protein-nucleic acid complexes". In: *Curr. Opin. Struct. Biol.* 94, 103079. DOI: 10.1016/j.sbi.2025.103079.
- Darnell Jr, J. E. (2002). "Transcription factors as targets for cancer therapy". In: *Nat. Rev. Cancer* 2.10, pp. 740–749. DOI: 10.1038/nrc906.
- Darty, K., A. Denise, and Y. Ponty (2009). "VARNA: Interactive drawing and editing of the RNA secondary structure". In: *Bioinformatics* 25.15, pp. 1974–1975. DOI: 10.1093/bioinformatics/btp250.
- Davies, D. R., E. A. Padlan, and S. Sheriff (1990). "Antibody-antigen complexes". In: *Annu. Rev. Biochem.* 59, pp. 439–473. DOI: 10.1146/annurev.bi.59.070190.002255.
- De Las Rivas, J. and C. Fontanillo (2010). "Protein-protein interactions essentials: key concepts to building and analyzing interactome networks". In: *PLoS Comput. Biol.* 6.6, e1000807. DOI: 10.1371/journal.pcbi.1000807.
- Deeks, E. D. (2016). "Venetoclax: First global approval". In: *Drugs* 76.9, pp. 979–987. DOI: 10.1007/s40265-016-0596-x.

- Dhillon, S. (2023). “Adagrasib: First approval”. In: *Drugs* 83.3, pp. 275–285. DOI: 10.1007/s40265-023-01839-y.
- Dominguez, C., R. Boelens, and A. M. J. J. Bonvin (2003). “HADDOCK: a protein-protein docking approach based on biochemical or biophysical information”. In: *J. Am. Chem. Soc.* 125.7, pp. 1731–1737. DOI: 10.1021/ja026939x.
- Doudna, J. A. and T. J. Richmond (2001). “Protein–nucleic acid interactions”. In: *Curr. Opin. Struct. Biol.* 11, pp. 11–13. DOI: 10.1016/s0959-440x(00)00178-0.
- Draper, D. E. (1995). “Protein-RNA recognition”. In: *Annu. Rev. Biochem.* 64.1, pp. 593–620. DOI: 10.1146/annurev.bi.64.070195.003113.
- Dubach, V. R. A. and A. Guskov (2020). “The resolution in X-ray crystallography and single-particle cryogenic electron microscopy”. In: *Crystals (Basel)* 10.7, 580. DOI: 10.3390/cryst10070580.
- Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowell, and P. Schultz (1988). “Cryo-electron microscopy of vitrified specimens”. In: *Q. Rev. Biophys.* 21.2, pp. 129–228. DOI: 10.1017/s0033583500004297.
- Dunbar, J., K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, and C. M. Deane (2014). “SAbDab: the structural antibody database”. In: *Nucleic Acids Res.* 42.D1, pp. D1140–D1146. DOI: 10.1093/nar/gkt1043.
- Dunbrack Jr, R. L. (2025). “Rēs ipSAE loquunt: What’s wrong with AlphaFold’s ipTM score and how to fix it”. In: *bioRxiv*. DOI: 10.1101/2025.02.10.637595.
- Edmunds, N. S., S. M. A. Alharbi, A. G. Genc, R. Adiyaman, and L. J. McGuffin (2023). “Estimation of model accuracy in CASP15 using the ModFOLDdock server”. In: *Proteins* 91.12, pp. 1871–1878. DOI: 10.1002/prot.26532.
- Elhabashy, H., F. Merino, V. Alva, O. Kohlbacher, and A. N. Lupas (2022). “Exploring protein-protein interactions at the proteome level”. In: *Structure* 30.4, pp. 462–475. DOI: 10.1016/j.str.2022.02.004.
- Evans, R. et al. (2022). “Protein complex prediction with AlphaFold-Multimer”. In: *bioRxiv*. DOI: 10.1101/2021.10.04.463034.
- Eyrich, V. A., M. A. Martí-Renom, D. Przybylski, M. S. Madhusudhan, A. Fiser, F. Pazos, A. Valencia, A. Sali, and B. Rost (2001). “EVA: continuous automatic evaluation of protein structure prediction servers”. In: *Bioinformatics* 17.12, pp. 1242–1243. DOI: 10.1093/bioinformatics/17.12.1242.
- Fadini, A., G. Studer, and R. J. Read (2026). “Model quality assessment for CASP16”. In: *Proteins* 94.1, pp. 302–313. DOI: 10.1002/prot.70037.
- Federhen, S. (2012). “The NCBI taxonomy database”. In: *Nucleic Acids Res.* 40.D1, pp. D136–43. DOI: 10.1093/nar/gkr1178.
- Fenwick, R. B., S. Esteban-Martín, and X. Salvatella (2011). “Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles”. In: *Eur. Biophys. J.* 40.12, pp. 1339–1355. DOI: 10.1007/s00249-011-0754-8.
- Ferré-D’Amaré, A. R., K. Zhou, and J. A. Doudna (1998). “A general module for RNA crystallization”. In: *J. Mol. Biol.* 279.3, pp. 621–631. DOI: 10.1006/jmbi.1998.1789.
- Frank, J. (2002). “Single-particle imaging of macromolecules by cryo-electron microscopy”. In: *Annu. Rev. Biophys. Biomol. Struct.* 31, pp. 303–319. DOI: 10.1146/annurev.biophys.31.082901.134202.
- Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li (2012). “CD-HIT: accelerated for clustering the next-generation sequencing data”. In: *Bioinformatics* 28.23, pp. 3150–3152. DOI: 10.1093/bioinformatics/bts565.

- Gao, M. and J. Skolnick (2010). "Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected". In: *Proc. Natl. Acad. Sci. U. S. A.* 107.52, pp. 22517–22522. doi: 10.1073/pnas.1012820107.
- Genz, L. R., T. Mulvaney, S. Nair, and M. Topf (2023). "PICKLUSTER: a protein-interface clustering and analysis plug-in for UCSF ChimeraX". In: *Bioinformatics* 39.11, btad629. doi: 10.1093/bioinformatics/btad629.
- Genz, L. R., S. Nair, N. Nagar, and M. Topf (2025a). "Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions". In: *Protein Sci.* 34.11, e70327. doi: 10.1002/pro.70327.
- Genz, L. R., S. Nair, A. Sweeney, and M. Topf (2025b). "Drug targeting of protein-nucleic acid interactions". In: *Curr. Opin. Struct. Biol.* 95, 103165. doi: 10.1016/j.sbi.2025.103165.
- Genz, L. R. and M. Topf (2026). "ProNA3D: Distance-based analysis of nucleic acid-containing interfaces". In: *bioRxiv*. doi: 10.64898/2026.04.16.719043.
- George, R. A. and J. Heringa (2002). "An analysis of protein domain linkers: their classification and role in protein folding". In: *Protein Eng.* 15.11, pp. 871–879. doi: 10.1093/protein/15.11.871.
- Göbel, U., C. Sander, R. Schneider, and A. Valencia (1994). "Correlated mutations and residue contacts in proteins". In: *Proteins* 18.4, pp. 309–317. doi: 10.1002/prot.340180402.
- Goddard, T. D., C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin (2018). "UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System". In: *Protein Sci.* 27.1, pp. 14–25. doi: 10.1002/pro.3235.
- Gomez, P. G. M. and O. Kovalevskiy (2025a). *Confidence Scores in AlphaFold-Multimer*. EMBL-EBI Training. URL: <https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/confidence-scores-in-alphafold-multimer/> (visited on 04/15/2026).
- (2025b). *pLDDT: Understanding Local Confidence*. EMBL-EBI Training. URL: <https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/plddt-understanding-local-confidence/> (visited on 04/15/2026).
- Good, M. C., J. G. Zalatan, and W. A. Lim (2011). "Scaffold proteins: hubs for controlling the flow of cellular information". In: *Science* 332.6030, pp. 680–686. doi: 10.1126/science.1198701.
- Grey, J. L. and D. H. Thompson (2010). "Challenges and opportunities for new protein crystallization strategies in structure-based drug design". In: *Expert Opin. Drug Discov.* 5.11, pp. 1039–1045. doi: 10.1517/17460441.2010.515583.
- Guo, J. and H.-X. Zhou (2016). "Protein allostery and conformational dynamics". In: *Chem. Rev.* 116.11, pp. 6503–6515. doi: 10.1021/acs.chemrev.5b00590.
- Haga, C. L. and D. G. Phinney (2023). "Strategies for targeting RNA with small molecule drugs". In: *Expert Opin. Drug Discov.* 18.2, pp. 135–147. doi: 10.1080/17460441.2022.2111414.
- Hagberg, A., P. J. Swart, and D. A. Schult (2008). "Exploring network structure, dynamics, and function using NetworkX". In: Los Alamos National Laboratory (LANL).

- Hakem, R. (2008). "DNA-damage repair; the good, the bad, and the ugly". In: *EMBO J.* 27.4, pp. 589–605. doi: 10.1038/emboj.2008.15.
- Halbeisen, R. E., A. Galgano, T. Scherrer, and A. P. Gerber (2008). "Post-transcriptional gene regulation: from genome-wide studies to principles". In: *Cell. Mol. Life Sci.* 65.5, pp. 798–813. doi: 10.1007/s00018-007-7447-6.
- Heel, M. van and M. Schatz (2005). "Fourier shell correlation threshold criteria". In: *J. Struct. Biol.* 151.3, pp. 250–262. doi: 10.1016/j.jsb.2005.05.009.
- Heifetz, A., E. Katchalski-Katzir, and M. Eisenstein (2002). "Electrostatics in protein-protein docking". In: *Protein Sci.* 11.3, pp. 571–587. doi: 10.1110/ps.26002.
- Hekkelman, M. L., D. Á. Salmoral, A. Perrakis, and R. P. Joosten (2025). "DSSP 4: FAIR annotation of protein secondary structure". In: *Protein Sci.* 34.8, e70208. doi: 10.1002/pro.70208.
- Heming, J. D., J. B. Huffman, L. M. Jones, and F. L. Homa (2014). "Isolation and characterization of the herpes simplex virus 1 terminase complex". In: *J. Virol.* 88.1, pp. 225–236. doi: 10.1128/JVI.02632-13.
- Henderson, R. (2018). "From electron crystallography to single particle CryoEM (Nobel lecture)". In: *Angew. Chem. Int. Ed Engl.* 57.34, pp. 10804–10825. doi: 10.1002/anie.201802731.
- Hermann, T. and D. J. Patel (1999). "Stitching together RNA tertiary architectures". In: *J. Mol. Biol.* 294.4, pp. 829–849. doi: 10.1006/jmbi.1999.3312.
- Hertel, K. J. and B. R. Graveley (2005). "RS domains contact the pre-mRNA throughout spliceosome assembly". In: *Trends Biochem. Sci.* 30.3, pp. 115–118. doi: 10.1016/j.tibs.2005.01.002.
- Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling (2006). "Comparison of multiple Amber force fields and development of improved protein backbone parameters". In: *Proteins* 65.3, pp. 712–725. doi: 10.1002/prot.21123.
- Hubbard, R. E. and M. K. Haider (2010). *Hydrogen Bonds in Proteins: Role and Strength*. In: *eLS*. Chichester: John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0003011.pub2.
- Hudson, W. H. and E. A. Ortlund (2014). "The structure, function and evolution of proteins that bind DNA and RNA". In: *Nat. Rev. Mol. Cell Biol.* 15.11, pp. 749–760. doi: 10.1038/nrm3884.
- Humphrey, W., A. Dalke, and K. Schulten (1996). "VMD: visual molecular dynamics". In: *Journal of molecular graphics* 14.1, pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.
- Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment". In: *Comput. Sci. Eng.* 9.3, pp. 90–95. doi: 10.1109/mcse.2007.55.
- Janeway, C. A. J., P. Travers, M. Walport, and M. J. Shlomchik (2001). "The Interaction of the Antibody Molecule with Specific Antigen". In: *Immunobiology: The Immune System in Health and Disease*. 5th ed. New York: Garland Science. (Visited on 05/25/2026).
- Janin, J. (1995). "Protein-protein recognition". In: *Prog. Biophys. Mol. Biol.* 64.2-3, pp. 145–166. doi: 10.1016/s0079-6107(96)00001-6.
- Janin, J., S. Miller, and C. Chothia (1988). "Surface, subunit interfaces and interior of oligomeric proteins". In: *J. Mol. Biol.* 204, pp. 155–164. doi: 10.1016/0022-2836(88)90606-7.
- Janin, J. (2005). "Assessing predictions of protein-protein interaction: the CAPRI experiment". In: *Protein Sci.* 14.2, pp. 278–283. doi: 10.1110/ps.041081905.

- Janin, J., R. P. Bahadur, and P. Chakrabarti (2008). "Protein-protein interaction and quaternary structure". In: *Q. Rev. Biophys.* 41.2, pp. 133–180. DOI: 10.1017/S0033583508004708.
- Järvelin, A. I., M. Noerenberg, I. Davis, and A. Castello (2016). "The new (dis)order in RNA regulation". In: *Cell Commun. Signal.* 14, 9. DOI: 10.1186/s12964-016-0132-3.
- Jia, X., X. He, C. Huang, J. Li, Z. Dong, and K. Liu (2024). "Protein translation: biological processes and therapeutic strategies for human diseases". In: *Signal Transduct. Target. Ther.* 9, 44. DOI: 10.1038/s41392-024-01749-9.
- Johnson, L. S., S. R. Eddy, and E. Portugaly (2010). "Hidden Markov model speed heuristic and iterative HMM search procedure". In: *BMC Bioinformatics* 11, 431. DOI: 10.1186/1471-2105-11-431.
- Jones, S., D. T. Daley, N. M. Luscombe, H. M. Berman, and J. M. Thornton (2001). "Protein-RNA interactions: a structural analysis". In: *Nucleic Acids Res.* 29.4, pp. 943–954. DOI: 10.1093/nar/29.4.943.
- Jones, S., P. van Heyningen, H. M. Berman, and J. M. Thornton (1999). "Protein-DNA interactions: A structural analysis". In: *J. Mol. Biol.* 287.5, pp. 877–896. DOI: 10.1006/jmbi.1999.2659.
- Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions". In: *Proc. Natl. Acad. Sci. U. S. A.* 93.1, pp. 13–20. DOI: 10.1073/pnas.93.1.13.
- Jubb, H., T. L. Blundell, and D. B. Ascher (2015). "Flexibility and small pockets at protein-protein interfaces: New insights into druggability". In: *Prog. Biophys. Mol. Biol.* 119.1, pp. 2–9. DOI: 10.1016/j.pbiomolbio.2015.01.009.
- Jubb, H. C., A. P. Higuero, B. Ochoa-Montaña, W. R. Pitt, D. B. Ascher, and T. L. Blundell (2017). "Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures". In: *J. Mol. Biol.* 429.3, pp. 365–371. DOI: 10.1016/j.jmb.2016.12.004.
- Jumper, J. et al. (2021). "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596, pp. 583–589. DOI: 10.1038/s41586-021-03819-2.
- Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.12, pp. 2577–2637. DOI: 10.1002/bip.360221211.
- Kang, D., Y. Lee, and J.-S. Lee (2020). "RNA-binding proteins in cancer: Functional and therapeutic perspectives". In: *Cancers (Basel)* 12.9, 2699. DOI: 10.3390/cancers12092699.
- Kappel, K., K. Zhang, Z. Su, A. M. Watkins, W. Kladwang, S. Li, G. Pintilie, V. V. Topkar, R. Rangan, I. N. Zheludev, J. D. Yesselman, W. Chiu, and R. Das (2020). "Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures". In: *Nat. Methods* 17.7, pp. 699–707. DOI: 10.1038/s41592-020-0878-9.
- Kaufmann, K. W., G. H. Lemmon, S. L. Deluca, J. H. Sheehan, and J. Meiler (2010). "Practically useful: what the Rosetta protein modeling suite can do for you". In: *Biochemistry* 49.14, pp. 2987–2998. DOI: 10.1021/bi902153g.
- Ke, A. and J. A. Doudna (2004). "Crystallization of RNA and RNA-protein complexes". In: *Methods* 34.3, pp. 408–414. DOI: 10.1016/j.ymeth.2004.03.027.
- Kendrew, J., R. Dickerson, B. Strandberg, R. Hart, D. Davies, D. Phillips, and V. Shore (1960). "Structure of myoglobin". In: *Nature* 185.422, pp. 427–1960.
- Kerpedjiev, P., S. Hammer, and I. L. Hofacker (2015). "Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams". In: *Bioinformatics* 31.20, pp. 3377–3379. DOI: 10.1093/bioinformatics/btv372.

- Keskin, O., B. Ma, and R. Nussinov (2005). "Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues". In: *J. Mol. Biol.* 345.5, pp. 1281–1294. doi: 10.1016/j.jmb.2004.10.077.
- Khoroshkin, M. et al. (2024). "Systematic identification of post-transcriptional regulatory modules". In: *Nat. Commun.* 15, 7872. doi: 10.1038/s41467-024-52215-7.
- Kimanius, D., L. Dong, G. Sharov, T. Nakane, and S. H. W. Scheres (2021). "New tools for automated cryo-EM single-particle analysis in RELION-4.0". In: *Biochem. J.* 478.24, pp. 4169–4185. doi: 10.1042/BCJ20210708.
- Klumpp, K., R. W. Ruigrok, and F. Baudin (1997). "Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure". In: *EMBO J.* 16.6, pp. 1248–1257. doi: 10.1093/emboj/16.6.1248.
- Koes, D. R. and C. J. Camacho (2012). "Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure". In: *Bioinformatics* 28.6, pp. 784–791. doi: 10.1093/bioinformatics/btr717.
- Krasnov, V. P., V. L. Andronova, A. V. Belyavsky, S. S. Borisevich, G. A. Galegov, O. F. Kandarakov, D. A. Gruzdev, O. A. Vozdvizhenskaya, and G. L. Levit (2023). "Large subunit of the human herpes simplex virus terminase as a promising target in design of anti-Herpesvirus agents". In: *Molecules* 28.21, 7375. doi: 10.3390/molecules28217375.
- Kretsch, R. C., A. M. Hummer, S. He, R. Yuan, J. Zhang, T. Karagianes, Q. Cong, A. Kryshtafovych, and R. Das (2026). "Assessment of nucleic acid structure prediction in CASP16". In: *Proteins* 94.1, pp. 192–217. doi: 10.1002/prot.70072.
- Kretsch, R. C. et al. (2023). "RNA target highlights in CASP15: Evaluation of predicted models by structure providers". In: *Proteins* 91.12, pp. 1600–1615. doi: 10.1002/prot.26550.
- Krishnan, V. V. and B. Rupp (2012). "Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy". In: *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0002716.pub2.
- Krissinel, E. and K. Henrick (2007). "Inference of Macromolecular Assemblies from Crystalline State". In: *J. Mol. Biol.* 372.3, pp. 774–797. doi: 10.1016/j.jmb.2007.05.022.
- Krüger, D. M., S. Neubacher, and T. N. Grossmann (2018). "Protein-RNA interactions: structural characteristics and hotspot amino acids". In: *RNA* 24.11, pp. 1457–1465. doi: 10.1261/rna.066464.118.
- Kryshtafovych, A., A. Barbato, K. Fidelis, B. Monastyrskyy, T. Schwede, and A. Tramontano (2014). "Assessment of the assessment: evaluation of the model quality estimates in CASP10". In: *Proteins* 82.S2, pp. 112–126. doi: 10.1002/prot.24347.
- Kryshtafovych, A., T. Schwede, M. Topf, K. Fidelis, and J. Moult (2019). "Critical assessment of methods of protein structure prediction (CASP)-Round XIII". In: *Proteins* 87.12, pp. 1011–1020. doi: 10.1002/prot.25823.
- (2021). "Critical assessment of methods of protein structure prediction (CASP)-Round XIV". In: *Proteins* 89.12, pp. 1607–1617. doi: 10.1002/prot.26237.
- (2026). "Progress and bottlenecks for deep learning in computational structure biology: CASP round XVI". In: *Proteins* 94.1, pp. 5–14. doi: 10.1002/prot.70076.

- Lafita, A., S. Bliven, A. Kryshchak, M. Bertoni, B. Monastyrskyy, J. M. Duarte, T. Schwede, and G. Capitani (2018). "Assessment of protein assembly prediction in CASP12". In: *Proteins* 86.S1, pp. 247–256. doi: 10.1002/prot.25408.
- Lawrence, M. C. and P. M. Colman (1993). "Shape complementarity at protein/protein interfaces". In: *J. Mol. Biol.* 234.4, pp. 946–950. doi: 10.1006/jmbi.1993.1648.
- Le Gall, T., P. R. Romero, M. S. Cortese, V. N. Uversky, and A. K. Dunker (2007). "Intrinsic disorder in the Protein Data Bank". In: *J. Biomol. Struct. Dyn.* 24.4, pp. 325–342. doi: 10.1080/07391102.2007.10507123.
- Lensink, M. F., R. Méndez, and S. J. Wodak (2007). "Docking and scoring protein complexes: CAPRI 3rd Edition". In: *Proteins* 69.4, pp. 704–718. doi: 10.1002/prot.21804.
- Lensink, M. F. and S. J. Wodak (2013). "Docking, scoring, and affinity prediction in CAPRI". In: *Proteins* 81.12, pp. 2082–2095. doi: 10.1002/prot.24428.
- Leontis, N. B. and E. Westhof (2001). "Geometric nomenclature and classification of RNA base pairs". In: *RNA* 7.4, pp. 499–512. doi: 10.1017/s1355838201002515.
- Li, W., R. D. Schaeffer, Z. Otwinowski, and N. V. Grishin (2016). "Estimation of uncertainties in the Global Distance Test (GDT_TS) for CASP models". In: *PLoS One* 11.5, e0154786. doi: 10.1371/journal.pone.0154786.
- Li, Y., X. Zhang, and D. Cao (2013). "The role of shape complementarity in the protein-protein interactions". In: *Sci. Rep.* 3, 3271. doi: 10.1038/srep03271.
- Li, Z. and P. D. Nagy (2011). "Diverse roles of host RNA binding proteins in RNA virus replication". In: *RNA Biol.* 8.2, pp. 305–315. doi: 10.4161/rna.8.2.15391.
- Lieberman-Blum, S. S., H. B. Fung, and J. C. Bandres (2008). "Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection". In: *Clin. Ther.* 30.7, pp. 1228–1250. doi: 10.1016/s0149-2918(08)80048-3.
- Liljas, A., L. Liljas, J. Piskur, G. Lindblom, P. Nissen, and M. Kjeldgaard (2009). *Textbook of Structural Biology*. WORLD SCIENTIFIC. ISBN: 9789812772077.
- Lill, M. A. (2011). "Efficient incorporation of protein flexibility and dynamics into molecular docking simulations". In: *Biochemistry* 50.28, pp. 6157–6169. doi: 10.1021/bi2004558.
- Lin, P., H. Li, and S.-Y. Huang (2024). "Deep learning in modeling protein complex structures: From contact prediction to end-to-end approaches". In: *Curr. Opin. Struct. Biol.* 85, 102789. doi: 10.1016/j.sbi.2024.102789.
- Lin, Z., H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, and A. Rives (2023). "Evolutionary-scale prediction of atomic-level protein structure with a language model". In: *Science* 379.6637, pp. 1123–1130. doi: 10.1126/science.ade2574.
- Lipfert, J., G. M. Skinner, J. M. Keegstra, T. Hensgens, T. Jager, D. Dulin, M. Köber, Z. Yu, S. P. Donkers, F.-C. Chou, R. Das, and N. H. Dekker (2014). "Double-stranded RNA under force and torque: similarities to and striking differences from double-stranded DNA". In: *Proc. Natl. Acad. Sci. U. S. A.* 111.43, pp. 15408–15413. doi: 10.1073/pnas.1407197111.
- Liu, S., B. Li, Q. Liang, A. Liu, L. Qu, and J. Yang (2020). "Classification and function of RNA-protein interactions". In: *Wiley Interdiscip. Rev. RNA* 11.6, e1601. doi: 10.1002/wrna.1601.

- Lo Conte, L., C. Chothia, and J. Janin (1999). "The atomic structure of protein-protein recognition sites". In: *J. Mol. Biol.* 285.5, pp. 2177–2198. DOI: 10.1006/jmbi.1998.2439.
- London, N., B. Raveh, and O. Schueler-Furman (2013). "Druggable protein-protein interactions—from hot spots to hot segments". In: *Curr. Opin. Chem. Biol.* 17.6, pp. 952–959. DOI: 10.1016/j.cbpa.2013.10.011.
- Look, A. T. (1997). "Oncogenic transcription factors in the human acute leukemias". In: *Science* 278.5340, pp. 1059–1064. DOI: 10.1126/science.278.5340.1059.
- López-Muñoz, A. D. and J. W. Yewdell (2024). "Cell surface RNA virus nucleocapsid proteins: a viral strategy for immunosuppression?" In: *Npj Viruses* 2, 41. DOI: 10.1038/s44298-024-00051-3.
- Lorenz, R., S. H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker (2011). "ViennaRNA Package 2.0". In: *Algorithms Mol. Biol.* 6, 26. DOI: 10.1186/1748-7188-6-26.
- Lu, X.-J., H. J. Bussemaker, and W. K. Olson (2015). "DSSR: an integrated software tool for dissecting the spatial structure of RNA". In: *Nucleic Acids Res.* 43.21, e142. DOI: 10.1093/nar/gkv716.
- Lunde, B. M., C. Moore, and G. Varani (2007). "RNA-binding proteins: modular design for efficient function". In: *Nat. Rev. Mol. Cell Biol.* 8.6, pp. 479–490. DOI: 10.1038/nrm2178.
- Ma, H., X. Jia, K. Zhang, and Z. Su (2022). "Cryo-EM advances in RNA structure determination". In: *Signal Transduct. Target. Ther.* 7.1, p. 58. DOI: 10.1038/s41392-022-00916-0.
- Machulin, A. V., J. V. Badaeva, S. Y. Grishin, E. I. Deryusheva, and O. V. Galzitskaya (2026). "In silico ligand-based screening of PDB database for searching unique motifs against SARS-CoV-2". In: *Biomolecules* 16.1, 163. DOI: 10.3390/biom16010163.
- Mariani, V., M. Biasini, A. Barbato, and T. Schwede (2013). "IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests". In: *Bioinformatics* 29.21, pp. 2722–2728. DOI: 10.1093/bioinformatics/btt473.
- Marion, D. (2013). "An introduction to biological NMR spectroscopy". In: *Mol. Cell. Proteomics* 12.11, pp. 3006–3025. DOI: 10.1074/mcp.0113.030239.
- Martí-Renom, M. A., A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Sali (2000). "Comparative protein structure modeling of genes and genomes". In: *Annu. Rev. Biophys. Biomol. Struct.* 29, pp. 291–325. DOI: 10.1146/annurev.biophys.29.1.291.
- Martick, M., T.-S. Lee, D. M. York, and W. G. Scott (2008). "Solvent structure and hammerhead ribozyme catalysis". In: *Chem. Biol.* 15.4, pp. 332–342. DOI: 10.1016/j.chembiol.2008.03.010.
- Maurakis, S. A., A. C. O'Donnell, I. Botos, R. Ghirlando, B. W. Davies, and S. K. Buchanan (2025). "Structural insights into Cir-mediated killing by the antimicrobial protein Microcin V". In: *Commun. Biol.* 8, 1449. DOI: 10.1038/s42003-025-08846-7.
- McGuffin, L. J., S. N. Alhaddad, B. Behzadi, N. S. Edmunds, A. G. Genc, and R. Adiyaman (2025). "Prediction and quality assessment of protein quaternary structure models using the MultiFOLD2 and ModFOLDdock2 servers". In: *Nucleic Acids Res.* 53.W1, W472–W477. DOI: 10.1093/nar/gkaf336.
- Melo, F. and A. Sali (2007). "Fold assessment for comparative protein structure modeling". In: *Protein Sci.* 16.11, pp. 2412–2426. DOI: 10.1110/ps.072895107.

- Mendez, J. H., A. Mehrani, P. Randolph, and S. Stagg (2019). "Throughput and resolution with a next-generation direct electron detector". In: *IUCr* 6.Pt 6, pp. 1007–1013. doi: 10.1107/S2052252519012661.
- Méndez, R., R. Leplae, L. De Maria, and S. J. Wodak (2003). "Assessment of blind predictions of protein-protein interactions: Current status of docking methods". In: *Proteins: Structure, Function and Genetics* 52.1, pp. 51–67. doi: 10.1002/prot.10393.
- Mihel, J., M. Sikić, S. Tomić, B. Jeren, and K. Vlahovick (2008). "PSAIA - protein structure and interaction analyzer". In: *BMC Struct. Biol.* 8, 21. doi: 10.1186/1472-6807-8-21.
- Mintseris, J. and Z. Weng (2005). "Structure, function, and evolution of transient and obligate protein-protein interactions". In: *Proc. Natl. Acad. Sci. U. S. A.* 102.31, pp. 10930–10935. doi: 10.1073/pnas.0502667102.
- Mirabello, C. and B. Wallner (2024). "DockQ v2: improved automatic quality measure for protein multimers, nucleic acids, and small molecules". In: *Bioinformatics* 40.10, btae586. doi: 10.1093/bioinformatics/btae586.
- Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger (2022). "ColabFold: making protein folding accessible to all". In: *Nat. Methods* 19.6, pp. 679–682. doi: 10.1038/s41592-022-01488-1.
- Mitra, R., A. S. Cohen, and R. Rohs (2024). "RNAescape: geometric mapping and customizable visualization of RNA structure". In: *Nucleic Acids Res.* 52.W1, W354–W361. doi: 10.1093/nar/gkae269.
- Mitra, R., A. S. Cohen, J. M. Sagendorf, H. M. Berman, and R. Rohs (2025a). "DNAproDB: an updated database for the automated and interactive analysis of protein-DNA complexes". In: *Nucleic Acids Res.* 53.D1, pp. D396–D402. doi: 10.1093/nar/gkae970.
- Mitra, R., A. S. Cohen, W. Y. Tang, H. Hosseini, Y. Hong, H. M. Berman, and R. Rohs (2025b). "RNAproDB: A webserver and interactive database for analyzing protein-RNA interactions". In: *J. Mol. Biol.* 437.15, 169012. doi: 10.1016/j.jmb.2025.169012.
- Moal, I. H., M. Torchala, P. A. Bates, and J. Fernández-Recio (2013). "The scoring of poses in protein-protein docking: current capabilities and future directions". In: *BMC Bioinformatics* 14, 286. doi: 10.1186/1471-2105-14-286.
- Mohan, B., P. Khanbadi, W. T. Fang, and J. Randhawa (2025). "Steps towards generation of an anti-GFP nanobody Escherichia coli surface display system using the BrkA autotransporter system". In: *UJEMI* 30.
- Montelione, G. T. (2012). "The Protein Structure Initiative: achievements and visions for the future". In: *F1000 Biol. Rep.* 4, p. 7. doi: 10.3410/B4-7.
- Moult, J., J. T. Pedersen, R. Judson, and K. Fidelis (1995). "A large-scale experiment to assess protein structure prediction methods". In: *Proteins* 23.3, pp. ii–v. doi: 10.1002/prot.340230303.
- Mulvaney, T., A. Kryshchak, and M. Topf (2026). "Cryo-EM analysis in CASP16". In: *Proteins* 94.1, pp. 447–459. doi: 10.1002/prot.70099.
- Nada, H., Y. Choi, S. Kim, K. S. Jeong, N. A. Meanwell, and K. Lee (2024). "New insights into protein-protein interaction modulators in drug discovery and therapeutic advance". In: *Signal Transduct. Target. Ther.* 9, 341. doi: 10.1038/s41392-024-02036-3.
- Nadassy, K., S. J. Wodak, and J. Janin (1999). "Structural features of protein-nucleic acid recognition sites". In: *Biochemistry* 38.7, pp. 1999–2017. doi: 10.1021/bi982362d.

- Nagaswamy, U., M. Larios-Sanz, J. Hury, S. Collins, Z. Zhang, Q. Zhao, and G. E. Fox (2002). "NCIR: a database of non-canonical interactions in known RNA structures". In: *Nucleic Acids Res.* 30.1, pp. 395–397. DOI: 10.1093/nar/30.1.395.
- Namba, K. and F. Makino (2022). "Recent progress and future perspective of electron cryomicroscopy for structural life sciences". In: *Microscopy (Oxf.)* 71.Supplement_1, pp. i3–i14. DOI: 10.1093/jmicro/dfab049.
- Neely, R. K., D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas, and A. C. Jones (2005). "Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes". In: *Nucleic Acids Res.* 33.22, pp. 6953–6960. DOI: 10.1093/nar/gki995.
- Neidle, S. (2021). "Beyond the double helix: DNA structural diversity and the PDB". In: *J. Biol. Chem.* 296, 100553. DOI: 10.1016/j.jbc.2021.100553.
- Nero, T. L., C. J. Morton, J. K. Holien, J. Wielens, and M. W. Parker (2014). "Oncogenic protein interfaces: small molecules, big challenges". In: *Nat. Rev. Cancer* 14.4, pp. 248–262. DOI: 10.1038/nrc3690.
- Noller, H. F. (2005). "RNA structure: reading the ribosome". In: *Science* 309.5740, pp. 1508–1514. DOI: 10.1126/science.1111771.
- Nooren, I. M. A. and J. M. Thornton (2003a). "Diversity of protein-protein interactions". In: *EMBO J.* 22.14, pp. 3486–3492. DOI: 10.1093/emboj/cdg359.
- (2003b). "Structural characterisation and functional significance of transient protein-protein interactions". In: *J. Mol. Biol.* 325.5, pp. 991–1018. DOI: 10.1016/S0022-2836(02)01281-0.
- Noy, A., A. Pérez, M. Márquez, F. J. Luque, and M. Orozco (2005). "Structure, recognition properties, and flexibility of the DNA.RNA hybrid". In: *J. Am. Chem. Soc.* 127.13, pp. 4910–4920. DOI: 10.1021/ja043293v.
- Nurullina, L. I., A. D. Biktimirov, M. M. Yusupov, and K. S. Usachev (2025). "Evolution of ribosome X-ray crystallography: from first crystals to atomic resolution". In: *Biophys. Rev.* 17.5, pp. 1199–1213. DOI: 10.1007/s12551-025-01357-z.
- O'Donoghue, S. I., D. S. Goodsell, A. S. Frangakis, F. Jossinet, R. A. Laskowski, M. Nilges, H. R. Saibil, A. Schafferhans, R. C. Wade, E. Westhof, and A. J. Olson (2010). "Visualization of macromolecular structures". In: *Nat. Methods* 7, S42–55. DOI: 10.1038/nmeth.1427.
- Ofran, Y. and B. Rost (2003). "Analysing six types of protein-protein interfaces". In: *J. Mol. Biol.* 325.2, pp. 377–387. DOI: 10.1016/S0022-2836(02)01223-8.
- Olechnovič, K. and Č. Venclovas (2019). "VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes". In: *Nucleic Acids Res.* 47.W1, W437–W442. DOI: 10.1093/nar/gkz367.
- (2021). "VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures". In: *Bioinformatics* 37.24, pp. 4873–4875. DOI: 10.1093/bioinformatics/btab448.
- (2023). "VoroIF-GNN: Voronoi tessellation-derived protein-protein interface assessment using a graph neural network". In: *Proteins* 91.12, pp. 1879–1888. DOI: 10.1002/prot.26554.
- Olson, W. K., S. Li, T. Kaukonen, A. V. Colasanti, Y. Xin, and X.-J. Lu (2019). "Effects of noncanonical base pairing on RNA folding: Structural context and spatial arrangements of G-A pairs". In: *Biochemistry* 58.20, pp. 2474–2487. DOI: 10.1021/acs.biochem.9b00122.

- Pal, A., P. Chakrabarti, and S. Dey (2022). "ProDFace: A web-tool for the dissection of protein-DNA interfaces". In: *Front. Mol. Biosci.* 9, 978310. DOI: 10.3389/fmolb.2022.978310.
- Pan, Y., Z. Wang, W. Zhan, and L. Deng (2018). "Computational identification of binding energy hot spots in protein-RNA complexes using an ensemble approach". In: *Bioinformatics* 34.9, pp. 1473–1480. DOI: 10.1093/bioinformatics/btx822.
- Pan, Y., S. Zhou, and J. Guan (2020). "Computationally identifying hot spots in protein-DNA binding interfaces using an ensemble approach". In: *BMC Bioinformatics* 21, 384. DOI: 10.1186/s12859-020-03675-3.
- PDB Statistics (2026). RCSB Protein Data Bank. URL: <https://www.rcsb.org/stats> (visited on 04/15/2026).
- Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12, pp. 2825–2830. DOI: 10.5555/1953048.2078195. eprint: 1201.0490.
- Peng, C., W. Ni, Q. Liu, G. Hu, and W. Zheng (2025). "A comprehensive benchmarking of the AlphaFold3 for predicting biomacromolecules and their interactions". In: *Brief. Bioinform.* 26.6, bbaf616. DOI: 10.1093/bib/bbaf616.
- Perkins, J. R., I. Diboun, B. H. Dessailly, J. G. Lees, and C. Orengo (2010). "Transient protein-protein interactions: structural, functional, and network properties". In: *Structure* 18.10, pp. 1233–1243. DOI: 10.1016/j.str.2010.08.007.
- Perutz, M. F., M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, and A. C. North (1960). "Structure of hæmoglobin: A three-dimensional Fourier synthesis at 5.5-Å. Resolution, obtained by X-ray analysis". In: *Nature* 185.4711, pp. 416–422. DOI: 10.1038/185416a0.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin (2004). "UCSF Chimera - A visualization system for exploratory research and analysis". In: *J. Comput. Chem.* 25.13, pp. 1605–1612. DOI: 10.1002/jcc.20084.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin (2021). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers". In: *Protein Sci.* 30.1, pp. 70–82. DOI: 10.1002/pro.3943.
- Pierce, B. G., K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, and Z. Weng (2014). "ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers". In: *Bioinformatics* 30.12, pp. 1771–1773. DOI: 10.1093/bioinformatics/btu097.
- Ponstingl, H., T. Kabir, D. Gorse, and J. M. Thornton (2005). "Morphological aspects of oligomeric protein structures". In: *Prog. Biophys. Mol. Biol.* 89.1, pp. 9–35. DOI: 10.1016/j.pbiomolbio.2004.07.010.
- Powers, D. M. W. (2020). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". In: *arXiv*. DOI: 10.48550/arXiv.2010.16061. arXiv: 2010.16061 [cs.LG].
- Ptashne, M. (2005). "Regulation of transcription: from lambda to eukaryotes". In: *Trends Biochem. Sci.* 30.6, pp. 275–279. DOI: 10.1016/j.tibs.2005.04.003.
- Punga, T., M. Larsson, E. Mujica, F. Rabelo Melo, C. Metzendorf, H. Zhang, D. Wang, M. den Hoed, L. Andersson, and D. Parrine (2026). "The C16orf87 protein is a subunit of the MIER corepressor complex controlling embryonic development and cell migration". In: *Sci. Rep.* 16, 13907. DOI: 10.1038/s41598-026-50740-7.
- Qin, H., H. Ni, Y. Liu, Y. Yuan, T. Xi, X. Li, and L. Zheng (2020). "RNA-binding proteins in tumor progression". In: *J. Hematol. Oncol.* 13, 90. DOI: 10.1186/s13045-020-00927-w.

- Qiu, L., Y. Yan, Z. Sun, J. Song, and J. Z. H. Zhang (2018). "Interaction entropy for computational alanine scanning in protein-protein binding". In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8.2, e1342. DOI: 10.1002/wcms.1342.
- Radivojac, P., Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, and A. K. Dunker (2004). "Protein flexibility and intrinsic disorder". In: *Protein Sci.* 13.1, pp. 71–80. DOI: 10.1110/ps.03128904.
- Reddy Chichili, V. P., V. Kumar, and J. Sivaraman (2013). "Linkers in the structural biology of protein-protein interactions: Linkers in Structural Biology". In: *Protein Sci.* 22.2, pp. 153–167. DOI: 10.1002/pro.2206.
- Reguei, A. and S. Murail (2025). "Af-analysis: a Python package for AlphaFold analysis". In: *J. Open Source Softw.* 10.107, 7577. DOI: 10.21105/joss.07577.
- Remmert, M., A. Biegert, A. Hauser, and J. Söding (2012). "HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment". In: *Nature methods* 9.2, pp. 173–175. DOI: 10.1038/nmeth.1818.
- Rhodes, G. (2010). *Crystallography made crystal clear: a guide for users of macromolecular models*. Elsevier. ISBN: 0125870736.
- Ritchie, D. W. (2008). "Recent progress and future directions in protein-protein docking". In: *Curr. Protein Pept. Sci.* 9.1, pp. 1–15. DOI: 10.2174/138920308783565741.
- Robin, X., P. Škrinjar, A. M. Waterhouse, G. Studer, G. Tauriello, J. Durairaj, and T. Schwede (2026). "Beyond single chains: Benchmarking macromolecular complex prediction methods with the Continuous Automated Model EvaluatiOn (CAMEO)". In: *Proteins* 94.1, pp. 403–413. DOI: 10.1002/prot.70060.
- Roddan, R., W. J. Nathan, J. A. Newman, A. El-Sagheer, D. M. Wilson, T. Brown, C. J. Schofield, and P. J. McHugh (2024). "Molecular insights into the stimulation of SNM1A nuclease activity by CSB during interstrand crosslink processing". In: *bioRxiv*. DOI: 10.1101/2024.09.05.611390.
- Rohl, C. A., C. E. M. Strauss, K. M. S. Misura, and D. Baker (2004). "Protein structure prediction using Rosetta". In: *Methods Enzymol.* 383, pp. 66–93. DOI: 10.1016/S0076-6879(04)83004-0.
- Rohs, R., X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann (2010). "Origins of specificity in protein-DNA recognition". In: *Annu. Rev. Biochem.* 79, pp. 233–269. DOI: 10.1146/annurev-biochem-060408-091030.
- Rosenhouse-Dantsker, A. (2026). *Unlocking protein structure-function relationships: Computational and experimental approaches and applications*. Ed. by A. Rosenhouse-Dantsker. Advances in Experimental Medicine and Biology. Cham: Springer Nature Switzerland. DOI: 10.1007/978-3-032-07511-6.
- Rosenthal, P. B. and R. Henderson (2003). "Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy". In: *J. Mol. Biol.* 333.4, pp. 721–745. DOI: 10.1016/j.jmb.2003.07.013.
- Ryan, D. P. and J. M. Matthews (2005). "Protein-protein interactions in human disease". In: *Curr. Opin. Struct. Biol.* 15.4, pp. 441–446. DOI: 10.1016/j.sbi.2005.06.001.
- Saenger, W. (1988). *Principles of nucleic acid structure*. 1st ed. Springer Advanced Texts in Chemistry. New York, NY: Springer. DOI: 10.1007/978-1-4612-5190-3.
- Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints". In: *J. Mol. Biol.* 234.3, pp. 779–815. DOI: 10.1006/jmbi.1993.1626.
- Sayle, R. A. and E. J. Milner-White (1995). "RASMOL: biomolecular graphics for all". In: *Trends Biochem. Sci.* 20.9, 374. DOI: 10.1016/S0968-0004(00)89080-5.

- Schake, P., S. N. Bolz, K. Linnemann, and M. Schroeder (2025). "PLIP 2025: introducing protein-protein interactions to the protein-ligand interaction profiler". In: *Nucleic Acids Res.* 53.W1, W463–W465. DOI: 10.1093/nar/gkaf361.
- Schargel, R. D., M. Z. Qayyum, A. S. Tanwar, R. C. Kalathur, and E. H. Kellogg (2025). "Structure of Fanzor2 reveals insights into the evolution of the TnpB superfamily". In: *Nat. Struct. Mol. Biol.* 32.2, pp. 243–246. DOI: 10.1038/s41594-024-01394-4.
- Schneider, B., J. C. Gelly, A. G. de Brevern, and J. Černý (2014). "Local dynamics of proteins and DNA evaluated from crystallographic B factors". In: *Acta Crystallogr. D Biol. Crystallogr.* 70.Pt 9, pp. 2413–2419. DOI: 10.1107/S1399004714014631.
- Schrödinger, LLC (2015). "The PyMOL Molecular Graphics System, Version 1.8".
- Schweke, H. et al. (2023). "Discriminating physiological from non-physiological interfaces in structures of protein complexes: A community-wide study". In: *Proteomics* 23.17, e2200323. DOI: 10.1002/pmic.202200323.
- Seabold, S. and J. Perktold (2010). "Statsmodels: Econometric and Statistical Modeling with Python". In: *Proceedings of the 9th Python in Science Conference*, pp. 57–61. DOI: 10.25080/Majora-92bf1922-011.
- Sehna, D., S. Bittrich, M. Deshpande, R. Svobodová, K. Berka, V. Bazgier, S. Velankar, S. K. Burley, J. Koča, and A. S. Rose (2021). "Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures". In: *Nucleic Acids Res.* 49.W1, W431–W437. DOI: 10.1093/nar/gkab314.
- Senior, A. W. et al. (2019). "Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13)". In: *Proteins* 87.12, pp. 1141–1148. DOI: 10.1002/prot.25834.
- (2020). "Improved protein structure prediction using potentials from deep learning". In: *Nature* 577.7792, pp. 706–710. DOI: 10.1038/s41586-019-1923-7.
- Sheinerman, F. B., R. Norel, and B. Honig (2000). "Electrostatic aspects of protein–protein interactions". In: *Current opinion in structural biology* 10.2, pp. 153–159. DOI: 10.1016/S0959-440X(00)00065-8.
- Simon, A. E. and L. Gehrke (2009). "RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs". In: *Biochim. Biophys. Acta* 1789.9-10, pp. 571–583. DOI: 10.1016/j.bbagr.2009.05.005.
- Smyth, M. S. and J. H. Martin (2000). "X ray crystallography". In: *Mol. Pathol.* 53.1, pp. 8–14. DOI: 10.1136/mp.53.1.8.
- Srivastava, A., T. Nagai, A. Srivastava, O. Miyashita, and F. Tama (2018). "Role of computational methods in going beyond X-ray crystallography to explore protein structure and dynamics". In: *Int. J. Mol. Sci.* 19.11, 3401. DOI: 10.3390/ijms19113401.
- Staley, J. P. and J. L. Woolford Jr (2009). "Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines". In: *Curr. Opin. Cell Biol.* 21.1, pp. 109–118. DOI: 10.1016/j.ceb.2009.01.003.
- Steinegger, M. and J. Söding (2017). "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets". In: *Nat. Biotechnol.* 35.11, pp. 1026–1028. DOI: 10.1038/nbt.3988.
- Studer, G., G. Tauriello, and T. Schwede (2023). "Assessment of the assessment-All about complexes". In: *Proteins* 91.12, pp. 1850–1860. DOI: 10.1002/prot.26612.
- Stumpf, M. P. H., T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf (2008). "Estimating the size of the human interactome". In: *Proc. Natl. Acad. Sci. U. S. A.* 105.19, pp. 6959–6964. DOI: 10.1073/pnas.0708078105.

- Suskiewicz, M. J. (2024). "The logic of protein post-translational modifications (PTMs): Chemistry, mechanisms and evolution of protein regulation through covalent attachments". In: *Bioessays* 46.3, e2300178. DOI: 10.1002/bies.202300178.
- Tao, Y. J. and Q. Ye (2010). "RNA virus replication complexes". In: *PLoS Pathog.* 6.7, e1000943. DOI: 10.1371/journal.ppat.1000943.
- Teufel, F., J. C. Refsgaard, M. A. Kasimova, K. Deibler, C. T. Madsen, C. Stahlhut, M. Grønberg, O. Winther, and D. Madsen (2023). "Deorphanizing peptides using structure prediction". In: *J. Chem. Inf. Model.* 63.9, pp. 2651–2655. DOI: 10.1021/acs.jcim.3c00378.
- Thachuk, C., A. Shmygelska, and H. H. Hoos (2007). "A replica exchange Monte Carlo algorithm for protein folding in the HP model". In: *BMC Bioinformatics* 8, 342. DOI: 10.1186/1471-2105-8-342.
- The Nobel Prize (2024). *The Nobel Prize in Chemistry 2024*. URL: <https://www.nobelprize.org/prizes/chemistry/2024/summary/> (visited on 04/15/2026).
- Thompson, M. C., T. O. Yeates, and J. A. Rodriguez (2020). "Advances in methods for atomic resolution macromolecular structure determination". In: *F1000Res.* 9, F1000 Faculty Rev-667. DOI: 10.12688/f1000research.25097.1.
- Tjong, V., L. Tang, S. Zauscher, and A. Chilkoti (2014). "'Smart' DNA interfaces". In: *Chem. Soc. Rev.* 43.5, pp. 1612–1626. DOI: 10.1039/c3cs60331h.
- Tripathy, U., ed. (2025). *Biochemical and biophysical methods in molecular and cellular biology*. Cham, Switzerland: Springer Nature. DOI: 10.1007/978-981-96-2088-3.
- Trueblood, K. N., H. B. Bürgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Shmueli, and S. C. Abrahams (1996). "Atomic displacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature". In: *Acta Crystallogr. A* 52.5, pp. 770–781. DOI: 10.1107/s0108767396005697.
- Tuffery, P. and P. Derreumaux (2012). "Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches". In: *J. R. Soc. Interface* 9.66, pp. 20–33. DOI: 10.1098/rsif.2011.0584.
- Tuteja, N. and R. Tuteja (2001). "Unraveling DNA repair in human: molecular mechanisms and consequences of repair defect". In: *Crit. Rev. Biochem. Mol. Biol.* 36.3, pp. 261–290. DOI: 10.1080/20014091074192.
- Vangone, A., R. Spinelli, V. Scarano, L. Cavallo, and R. Oliva (2011). "COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes". In: *Bioinformatics* 27.20, pp. 2915–2916. DOI: 10.1093/bioinformatics/btr484.
- Varadi, M., M. Tsenkov, and S. Velankar (2025). "Challenges in bridging the gap between protein structure prediction and functional interpretation". In: *Proteins* 93.1, pp. 400–410. DOI: 10.1002/prot.26614.
- Varadi, M. et al. (2022). "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models". In: *Nucleic Acids Res.* 50.D1, pp. D439–D444. DOI: 10.1093/nar/gkab1061.
- Verma, V. V., S. Vimal, M. K. Mishra, and V. K. Sharma (2025). "A comprehensive review on structural insights through molecular visualization: tools, applications, and limitations". In: *J. Mol. Model.* 31.6, 173. DOI: 10.1007/s00894-025-06402-y.
- Vetrano, P., K. Krall, L. Martinez, E. Traverso, T. Morosinotto, N. A. T. Irwin, Y. Mazor, and S. Ramundo (2026). "VIA1 is a conserved regulator of thylakoid membrane integrity that acts through VIPP1". In: *bioRxiv*. DOI: 10.64898/2026.03.26.714412.

- Virtanen, P. et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nat. Methods* 17.3, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.
- Voet, D., J. G. Voet, and C. W. Pratt (2016). *Fundamentals of biochemistry: Life at the molecular level*. 5th ed. Nashville, TN: John Wiley & Sons. ISBN: 9781118918401, 9781118918432.
- Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid". In: *Nature* 171.4356, pp. 737–738. DOI: 10.1038/171737a0.
- Wells, J. A. and C. L. McClendon (2007). "Reaching for high-hanging fruit in drug discovery at protein-protein interfaces". In: *Nature* 450.7172, pp. 1001–1009. DOI: 10.1038/nature06526.
- Wentinck, K., C. Gogou, and D. H. Meijer (2022). "Putting on molecular weight: Enabling cryo-EM structure determination of sub-100-kDa proteins". In: *Curr. Res. Struct. Biol.* 4, pp. 332–337. DOI: 10.1016/j.crstbi.2022.09.005.
- Westermarck, J., J. Ivaska, and G. L. Corthals (2013). "Identification of protein interactions involved in cellular signaling". In: *Mol. Cell. Proteomics* 12.7, pp. 1752–1763. DOI: 10.1074/mcp.R113.027771.
- Winter, A., A. P. Higueruelo, M. Marsh, A. Sigurdardottir, W. R. Pitt, and T. L. Blundell (2012). "Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery". In: *Q. Rev. Biophys.* 45.4, pp. 383–426. DOI: 10.1017/S0033583512000108.
- Wodak, S. J. and J. Janin (1978). "Computer analysis of protein-protein interaction". In: *J. Mol. Biol.* 124.2, pp. 323–342. DOI: 10.1016/0022-2836(78)90302-9.
- Wojdyr, M. (2022). "GEMMI: A library for structural biology". In: *J. Open Source Softw.* 7.73, 4200. DOI: 10.21105/joss.04200.
- wwPDB Consortium (2024). "EMDB-the Electron Microscopy Data Bank". In: *Nucleic Acids Res.* 52.D1, pp. D456–D465. DOI: 10.1093/nar/gkad1019.
- Xu, D. and Y. Zhang (2012). "Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field: QUARK Ab initio Prediction Method". In: *Proteins* 80.7, pp. 1715–1735. DOI: 10.1002/prot.24065.
- Xue, L. C., J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin, and A. Vangone (2016). "PRODIGY: a web server for predicting the binding affinity of protein-protein complexes". In: *Bioinformatics* 32.23, pp. 3676–3678. DOI: 10.1093/bioinformatics/btw514.
- Yan, C., F. Wu, R. L. Jernigan, D. Dobbs, and V. Honavar (2008). "Characterization of Protein-Protein Interfaces". In: *Protein J.* 27, pp. 59–70. DOI: 10.1007/s10930-007-9108-x.
- Yan, Y., H. Tao, J. He, and S.-Y. Huang (2020). "The HDock server for integrated protein-protein docking". In: *Nat. Protoc.* 15.5, pp. 1829–1852. DOI: 10.1038/s41596-020-0312-x.
- Yang, H., F. Jossinet, N. Leontis, L. Chen, J. Westbrook, H. Berman, and E. Westhof (2003). "Tools for the automatic identification and classification of RNA base pairs". In: *Nucleic Acids Res.* 31.13, pp. 3450–3460. DOI: 10.1093/nar/gkg529.
- Yang, Y., P. Yang, N. Wang, Z. Chen, D. Su, Z. H. Zhou, Z. Rao, and X. Wang (2020). "Architecture of the herpesvirus genome-packaging complex and implications for DNA translocation". In: *Protein Cell* 11.5, pp. 339–351. DOI: 10.1007/s13238-020-00710-0.
- Yao, H., D. M. Kristensen, I. Mihalek, M. E. Sowa, C. Shaw, M. Kimmel, and Lichtarge (2003). "An accurate, sensitive, and scalable method to identify functional sites in

- protein structures". In: *Journal of molecular biology* 326.1, pp. 255–261. DOI: 10.1016/S0022-2836(02)01336-0.
- Ye, J.-D., V. Tereshko, J. K. Frederiksen, A. Koide, F. A. Fellouse, S. S. Sidhu, S. Koide, A. A. Kossiakoff, and J. A. Piccirilli (2008). "Synthetic antibodies for specific recognition and crystallization of structured RNA". In: *Proc. Natl. Acad. Sci. U. S. A.* 105.1, pp. 82–87. DOI: 10.1073/pnas.0709082105.
- Yesudhas, D., M. Batool, M. A. Anwar, S. Panneerselvam, and S. Choi (2017). "Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors". In: *Genes (Basel)* 8.8, p. 192. DOI: 10.3390/genes8080192.
- Yin, R., B. Y. Feng, A. Varshney, and B. G. Pierce (2022). "Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants". In: *Protein Sci.* 31.8, e4379. DOI: 10.1002/pro.4379.
- Yip, K. M., N. Fischer, E. Paknia, A. Chari, and H. Stark (2020). "Atomic-resolution protein structure determination by cryo-EM". In: *Nature* 587.7832, pp. 157–161. DOI: 10.1038/s41586-020-2833-4.
- You, L., J. Ma, J. Wang, D. Artamonova, M. Wang, L. Liu, H. Xiang, K. Severinov, X. Zhang, and Y. Wang (2019). "Structure studies of the CRISPR-Csm complex reveal mechanism of co-transcriptional interference". In: *Cell* 176, 239–253.e16. DOI: 10.1016/j.cell.2018.10.052.
- Zemla, A. (2003). "LGA: A method for finding 3D similarities in protein structures". In: *Nucleic Acids Res.* 31.13, pp. 3370–3374. DOI: 10.1093/nar/gkg571.
- Zhang, Q., Y. Gao, M. L. Baker, S. Liu, X. Jia, H. Xu, J. He, J. T. Kaelber, S. Weng, and W. Jiang (2023). "The structure of a 12-segmented dsRNA reovirus: New insights into capsid stabilization and organization". In: *PLoS Pathog.* 19.4, e1011341. DOI: 10.1371/journal.ppat.1011341.
- Zhang, Q., M. Sanner, and A. J. Olson (2009). "Shape complementarity of protein-protein complexes at multiple resolutions". In: *Proteins* 75.2, pp. 453–467. DOI: 10.1002/prot.22256.
- Zhang, W., D. Xiao, Q. Mao, and H. Xia (2023). "Role of neuroinflammation in neurodegeneration development". In: *Signal Transduct. Target. Ther.* 8, 267. DOI: 10.1038/s41392-023-01486-5.
- Zhang, Y. and J. Skolnick (2004). "Scoring function for automated assessment of protein structure template quality". In: *Proteins* 57.4, pp. 702–710. DOI: 10.1002/prot.20264.
- Zhou, X., F. Huang, J. Zhang, S. Gong, L. Wei, F. Liu, and Z. Song (2025). "The immunomodulatory effects of sodium new houttuylfonate on different states of macrophage against *Aspergillus fumigatus* infection via distinct mechanism in invasive pulmonary aspergillosis". In: *Chin. Med.* 20, 102. DOI: 10.1186/s13020-025-01157-3.
- Zhu, W., A. Shenoy, P. Kundrotas, and A. Elofsson (2023). "Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes". In: *Bioinformatics* 39.7, btad424. DOI: 10.1093/bioinformatics/btad424.

Acknowledgements

This thesis would not have been possible without the support of many people to whom I am deeply grateful.

First and foremost, I would like to thank Maya for giving me the opportunity to work on this project and for creating such a supportive research environment. It has been a pleasure working with you ever since I joined the group during my first internship as a Master's student, and I have greatly enjoyed our scientific discussions throughout the years. Thank you not only for your scientific guidance and mentorship, but also for encouraging me to take on new challenges and opportunities.

I would also like to thank Kay for taking on the role of co-supervisor and for always providing valuable feedback and support.

I am deeply grateful to all current and former members of the Topf group. Thank you for the countless scientific discussions, for your help and advice, and for making the group such a welcoming and enjoyable place to work.

A special thank you goes to Aaron for introducing me to the world of computational biology at a time when I was still convinced that my future would be entirely in the wet lab. Thank you, Tom, for patiently answering my questions whenever I appeared from behind the shelf with "Hey Tom, can I ask you something?". Thank you, Matthias, for the dad jokes and for consistently coming up with excellent tool names. Thank you, Fabian, for being the fastest PhD thesis proofreader imaginable. Thank you, Gabriele, for always checking in and for your support. Gwen, thank you for your energy and humor, and for being the best hype woman anyone could ask for. And finally, thank you, Sanjana, my conference roomie, for all your help, advice, and patience during the countless hours you had to sit through my meetings.

I would also like to thank my family and friends for accompanying me throughout this journey, celebrating the successes and helping me through the more difficult moments. A very special thank you goes to my parents, who always believed in me and gave me every opportunity I could have wished for.

And finally, thank you, Vincent. I truly do not know how I could have done this without you.

Affidavit

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated. If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

26.05.2026

Date



Signature

(Luca Ragna Genz)